THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

Section B

VOL. 68, PART 1 1 January 1955 No. 421 B

An Electrostatic Problem Involving a Non-Linear Fluid Dielectric

BY R. CADE
University College of the West Indies, Jamaica

Communicated by R. Fürth; MS. received 28th September 1954

Abstract. An approximate calculation is made of the electrostatic couple on an anisotropic dielectric sphere influenced by a uniform field and immersed in a slightly non-linear fluid dielectric. The non-linearity is contained in a term in the dielectric constant proportional to the square of the field strength, and may arise from an intrinsic property of the dielectric or through electrostriction. These alternatives are discussed in the light of numerical values for the couple, and the possibility of experimental application is considered. The problem has paramagnetic and steady-current analogues, and the results are adapted to these problems.

§ 1. INTRODUCTION

CONSIDER an anisotropic non-conducting sphere influenced by a uniform electric field and immersed in a non-conducting fluid whose dielectric constant is given by

$$\epsilon=\epsilon_0(1+\nu E^2), \qquad \ldots\ldots(1)$$

where $\mathbf{E}$ is the electric intensity and ϵ_0 and ν are constants, the latter small. We shall calculate, to the first order in ν, the couple on the sphere due to the applied field. The problem for the special case of a linear dielectric ($\nu=0$) has been solved previously by the author (1953). The present problem is intrinsically interesting as an example of a non-linear electrostatic problem which has not so much symmetry as to make it more or less trivial. Moreover, an anisotropic sphere seems to be the *only* case of an immersed body for which the modification of the couple due to the non-linearity is mathematically tractable. The existence of this modification lends the possibility of experimental application, which will be discussed later in relation to the physical occurrence of dielectrics which obey equation (1).

By simple transformations, the results will be adapted to the analogous paramagnetic and steady-current problems.

§ 2. CALCULATION OF THE FIELD

We assume provisionally that the anisotropic solid dielectric is linear and perfectly rigid, so that the dielectric tensor $\epsilon_{\mu\nu}$ is field-independent. Call the radius of the sphere a. Choose the centre of the sphere as the origin of Cartesian coordinates and the dielectric axes through the centre as left-hand coordinate axes. Define spherical polar coordinates (r, θ, ϕ) by

$$x=r\cos\theta, \quad y=r\sin\theta\cos\phi, \quad z=r\sin\theta\sin\phi. \qquad \ldots\ldots(2)$$

We shall suppose that the direction of the original uniform field, of strength $\mathbf{F}$, is parallel to the xy plane, making an angle γ say with the x-direction. The loss of generality through this special choice of the $\mathbf{F}$-direction is trivial as regards the issues of mathematical interest and experimental application, but the saving of mathematical labour is great.

It follows from our choice of axes that the dielectric tensor has the diagonal form

$$\epsilon_{\mu\nu}=\begin{bmatrix} \epsilon_x & 0 & 0 \\ & \epsilon_y & 0 \\ & & \epsilon_z \end{bmatrix}, \qquad \ldots\ldots(3)$$

the non-zero components being the principal values. We define the tensor

$$K_{\mu\nu}=\epsilon_{\mu\nu}/\epsilon_0, \qquad \ldots\ldots(4)$$

having the same principal axes as $\epsilon_{\mu\nu}$ and principal values K_x, K_y and K_z.

(i) *The External Field*

Outside the sphere the potential satisfies the equation

$$\operatorname{div}(\epsilon \operatorname{grad} V)=0. \qquad \ldots\ldots(5)$$

We treat this by a perturbation method, and attempt to calculate the second term in a series of the form

$$V=V_0+\nu v_1+\nu^2 v_2+\ldots, \qquad \ldots\ldots(6)$$

where V_0 is the value of V for the linear problem and is given by

$$V_0=-Fr\left\{\left[1-\left(\frac{K_x-1}{2+K_x}\right)\frac{a^3}{r^3}\right]\cos\gamma P_1(\cos\theta) + \left[1-\left(\frac{K_y-1}{2+K_y}\right)\frac{a^3}{r^3}\right]\sin\gamma P_1{}^1(\cos\theta)\cos\phi\right\} \qquad \ldots\ldots(7)$$

(Cade 1953). From equations (5) and (6), working only to the first power of ν and putting $\nu v_1=\lambda$, we obtain

$$\nabla^2\lambda-\nu(\mathbf{E}_0, \operatorname{grad} E_0{}^2)=0 \qquad \ldots\ldots(8)$$

where

$$\mathbf{E}_0=-\operatorname{grad} V_0. \qquad \ldots\ldots(9)$$

Equation (8) can be solved by a method basically equivalent to that of the Legendre transform (Tranter 1950). When the second term is evaluated in terms of equation (7), we find that the result can be rearranged as a polynomial in the surface harmonics $P_1{}^m(\cos\theta)\cos m\phi$ $(m=0, 1)$, $P_3{}^m(\cos\theta)\cos m\phi$ $(m=0, \ldots 3)$. If we now express λ as a Laplace series,

$$\lambda=\sum_{n=0}^{\infty}\sum_{m=0}^{n}(\alpha_{nm}(r)\cos m\phi+\beta_{nm}(r)\sin m\phi)P_n{}^m(\cos\theta), \qquad \ldots\ldots(10)$$

the orthogonality of surface harmonics allows us to extract from (8) a set of ordinary differential equations for the α_{nm} and β_{nm}, which are as follows:

$$D^2\alpha_{nm}+D\alpha_{nm}+n(n+1)\alpha_{nm}=\nu F^3\left(f_{nm}\frac{a^3}{r^2}+g_{nm}\frac{a^6}{r^5}+h_{nm}\frac{a^9}{r^8}\right)$$

$$(n=1,3\,;\ m=0,\dots n), \qquad \dots\dots(11)$$

$$D^2\alpha_{nm}+D\alpha_{nm}+n(n+1)\alpha_{nm}=0 \quad (n=0,2,>3), \qquad \dots\dots(12)$$

$$D^2\beta_{nm}+D\beta_{nm}+n(n+1)\beta_{nm}=0 \quad (\text{all } n), \qquad \dots\dots(13)$$

where

$$D\equiv r\,d/dr \qquad \dots\dots(14)$$

and f_{nm}, g_{nm} and h_{nm} are known functions of the K's and γ.

The function Ar^n+B/r^{n+1}, where A and B are arbitrary constants, is the solution of all equations (12) and (13) and is the complementary function of (11). To satisfy conditions at infinity we must always have $A=0$. Guided by the simpler, linear problem, we shall tentatively take $B=0$ for equations (12) and (13), although not for (11), where we anticipate that the solution may be influenced by the boundary conditions. We then find, when the particular integrals of (11) are evaluated, that

$$\lambda=\nu F^3\left\{\sum_{m=0}^{1}\left(c_{1m}\frac{a^6}{r^5}+d_{1m}\frac{a^9}{r^8}+\frac{e_{1m}}{r^2}\right)P_1{}^m\cos m\phi\right.$$
$$\left.+\sum_{m=0}^{3}\left(b_{3m}\frac{a^3}{r^2}+c_{3m}\frac{a^6}{r^5}+d_{3m}\frac{a^9}{r^8}+\frac{e_{3m}}{r^4}\right)P_3{}^m\cos m\phi\right\}, \qquad \dots\dots(15)$$

where the b_{nm}, c_{nm} and d_{nm} are related to the f_{nm}, g_{nm} and h_{nm} and are known functions of the K's and γ, while the e_{nm} are the undetermined constants. We shall not quote the values of the b_{nm}, c_{nm} and d_{nm} here as the expressions are cumbersome and are of little interest in themselves.

(ii) *The Internal Field*

The potential inside the sphere satisfies the equation

$$\epsilon_x\frac{\partial^2U}{\partial x^2}+\epsilon_y\frac{\partial^2U}{\partial y^2}+\epsilon_z\frac{\partial^2U}{\partial z^2}=0, \qquad \dots\dots(16)$$

and the tentative step above can only be justified and the e_{nm} fixed if a solution of equation (16) can be found which, in conjunction with equation (15), satisfies the boundary conditions at the surface of the sphere. These conditions, derived from the respective physical assumptions that the boundary has no double layer† and no surface charge, are

$$V=U(r=a), \qquad \dots\dots(17)$$

$$\epsilon\frac{\partial V}{\partial r}=\overline{\left(\epsilon_x\frac{\partial U}{\partial x}\mathbf{i}+\epsilon_y\frac{\partial U}{\partial y}\mathbf{j}+\epsilon_z\frac{\partial U}{\partial z}\mathbf{k},\mathbf{n}\right)} \quad (r=a), \qquad \dots\dots(18)$$

where **i**, **j** and **k** are unit vectors in the respective Cartesian coordinate directions and **n** is a unit outward normal at the surface of the sphere.

The value of U in the linear case is

$$U_0=-3Fr\left\{\frac{\cos\gamma}{2+K_x}P_1(\cos\theta)+\frac{\sin\gamma}{2+K_y}P_1{}^1(\cos\theta)\cos\phi\right\} \qquad \dots\dots(19)$$

(Cade 1953), and our problem now is to add to U_0 a solution of equation (16), μ say, of order ν, such that, to the first power of ν, equations (17) and (18) can be satisfied for appropriate values of arbitrary constants.

---

† The significance of the double layer in the formulation of the boundary condition for the potential is frequently overlooked in purely 'classical' electrostatics.

By the transformation

$$\xi=\epsilon_x^{-1/2}x, \quad \eta=\epsilon_y^{-1/2}y, \quad \zeta=\epsilon_z^{-1/2}z, \qquad \ldots\ldots(20)$$

equation (16), now applied to μ, becomes Laplace's equation in the new variables:

$$\frac{\partial^2\mu}{\partial\xi^2}+\frac{\partial^2\mu}{\partial\eta^2}+\frac{\partial^2\mu}{\partial\zeta^2}=0. \qquad \ldots\ldots(21)$$

Formal solution of this equation could be carried out, but is rendered difficult by the fact that equation (20) transforms the spherical boundary to an ellipsoidal one, so that the solution is in ellipsoidal harmonics. A simpler and more elegant route to the solution is provided by the use of rational integral harmonics (Jeans 1943). Our guide in trying to find suitable ones is that, in the ultimate transformation to (r, θ, ϕ), the solution has, through equation (17), to be matched to equation (15), so that (*a*) we require six and no more independent ones, (*b*) in $\xi^p\eta^q\zeta^s$, $p+q+s=1$ or 3, and (*c*) no combination of ξ, η and ζ is allowable which leads to $\sin\phi$, $\sin 2\phi$ or $\sin 3\phi$. The six harmonics are easily found to be

$$\xi, \quad \eta, \quad \xi^3-3\xi\eta^2, \quad \eta^3-3\eta\xi^2, \quad \xi\zeta^2-\xi\eta^2, \quad \eta\zeta^2-\eta\xi^2, \qquad \ldots\ldots(22)$$

and we obtain μ by multiplying them by arbitrary constants and forming their sum. Clearly this solution satisfies the requirements of remaining finite at the origin. By equations (20) and (2) it can be transformed to (x, y, z) and (r, θ, ϕ) as required.

The application of equations (17) and (18), to the first order in ν, provides twelve linear equations in the six above-mentioned constants and the six e_{nm}. With their solution, therefore, the potential problem is solved. Again, we shall not quote the expressions for the constants as they are extremely cumbersome and are of little direct interest. We note in passing, however, that the evaluation of e_{10} and e_{11} is relatively simple, and these happen to be the only members of the set that are required in the calculation of the couple.

(iii) *Uniqueness of Solution*

That the solution in the case $\nu=0$ is the only solution is guaranteed by the uniqueness theorem.* If $\nu\neq0$, ϵ varies spatially. The uniqueness theorem still holds for ϵ a function of (x, y, z) provided the functional dependence is given,† but in the present case the situation is more subtle as the distribution of ϵ is not given but is determined by the field itself. With our perturbation procedure, however, the distribution at every stage in the perturbation series *is given in terms of the field as calculated at the preceding stage*, so that, by an inductive argument, the theorem is applicable provided the perturbation method is valid. The problem of uniqueness therefore belongs to the broader and more difficult question of the convergence of perturbation series, and we may conclude that provided our solution is correct, it is the only correct solution of this order.

---

* The field includes a region in which the dielectric is anisotropic, and the uniqueness theorem is not proved in the textbooks for anisotropic dielectrics. However, the generalization is easily carried out provided ϵ_x, ϵ_y and ϵ_z are positive quantities, which seems always to be the case in practice.

† Again, the theorem is not proved for non-uniform dielectrics. The generalization in this case is simple and offers no new features.

§ 3. CALCULATION OF THE COUPLE

The theory of mechanical action on immersed bodies, a long-controversial subject in electrostatic theory, has recently been re-examined and put on a rigorous footing by Brown (1951). The primary or direct electrical action on the body is given by Livens' stress tensor,†

$$R_{xx}=\frac{\epsilon E_x^2}{4\pi}-\frac{E^2}{8\pi},\quad R_{xy}=\frac{\epsilon E_x E_y}{4\pi}\quad \text{etc.,} \qquad \ldots\ldots(23)$$

to which must be added a tensor representing a secondary, hydrostatic effect due to the electrical forces in the fluid itself. The fact that there is no double layer means that certain amendments to Brown's theory proposed by the author (1954) have a null effect.

The secondary stress tensor depends critically upon the nature of the fluid dielectric, but under static conditions it is always a multiple of the fundamental second-rank tensor, that is, it can only provide normal stresses. By consideration of the forms that the secondary stress tensor can take under the different circumstances (to be discussed later) in which equation (1) is satisfied, the mechanical action on the sphere can, by symmetry, only be a couple. Normal stresses do not enter the calculation of the couple on a sphere, so that if we resolve vectors in the r, θ and ϕ directions, we only need the non-diagonal components of Livens' tensor which act on a surface-element facing the r-direction, that is, $\epsilon E_r E_\theta/4\pi$ and $\epsilon E_r E_\varphi/4\pi$.

The couple, which by symmetry is about the z-axis, is

$$G_z=-\frac{1}{4\pi}\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi}\epsilon E_r(E_\theta\cos\phi-E_\varphi\cos\theta\sin\phi)a^3\sin\theta\,d\phi\,d\theta, \qquad \ldots\ldots(24)$$

and evaluating the integral to the first power of ν, we obtain

$$\begin{aligned}G_z=\frac{3}{2}\epsilon_0F^2a^3\sin 2\gamma&\left(\frac{1}{2+K_x}-\frac{1}{2+K_y}\right)\\ \times\Bigg\{1+\frac{1}{15}\nu F^2&\Bigg[\frac{\cos^2\gamma}{2+K_x}\Bigg\{\frac{27(3K_x^2+2)(4-K_xK_y)}{(2+K_x)^2(2+K_y)}+\frac{3(13K_x^2+28K_x+22)}{2+K_x}\\ &-\frac{6(K_xK_y-K_x-K_y-8)}{2+K_y}-\frac{0(K_x-1)^2}{2+K_x}\left(\frac{K_x-1}{2+K_x}+\frac{K_y-1}{2+K_y}\right)\Bigg\}\\ &+\frac{\sin^2\gamma}{2+K_y}\Bigg\{\frac{27(3K_y^2+2)(4-K_yK_x)}{(2+K_y)^2(2+K_x)}+\frac{3(13K_y^2+28K_y+22)}{2+K_y}\\ &-\frac{6(K_yK_x-K_y-K_x-8)}{2+K_x}-\frac{20(K_y-1)^2}{2+K_y}\left(\frac{K_y-1}{2+K_y}+\frac{K_x-1}{2+K_x}\right)\Bigg\}\Bigg]\Bigg\}. \end{aligned} \qquad \ldots(25)$$

The first term in the outer curly brackets represents the couple previously obtained (Cade 1953) for the linear case,‡ while the remainder is the effect of the non-linearity.

§ 4. ANALOGOUS PROBLEMS

The problem has a paramagnetic analogue. Giving the symbols their corresponding paramagnetic significance, the whole of the preceding theory applies to this problem.

---

† Livens (1926) appears to have first given it, although he misunderstood its significance.

‡ The result in the paper cited has the opposite sign. This is due to the implicit use there of right-hand coordinates.

There is also a steady-current analogue. We suppose that, in addition to dielectric non-linearity, the fluid has conductive non-linearity, the conductivity obeying the equation

$$\sigma=\sigma_0(1+\nu' E^2), \qquad \ldots\ldots(26)$$

while the sphere is a linear anisotropic conductor with conductivity tensor $\sigma_{\mu\nu}$. The specification is the same as with the dielectric problem, except that we choose the coordinate axes to be parallel to the principal axes of $\sigma_{\mu\nu}$ rather than of $\epsilon_{\mu\nu}$, so that $\sigma_{\mu\nu}$ has the diagonal form

$$\sigma_{\mu\nu}=\begin{bmatrix} \sigma_x & 0 & 0 \\ & \sigma_y & 0 \\ & & \sigma_z \end{bmatrix}. \qquad \ldots\ldots\ldots(27)$$

We define the quantities

$$\tau_x=\sigma_x/\sigma_0, \quad \tau_y=\sigma_y/\sigma_0, \quad \tau_z=\sigma_z/\sigma_0, \qquad \ldots\ldots(28)$$

corresponding to the K_x, K_y and K_z that we have used previously.

Using the boundary condition (17) and the condition

$$\sigma\frac{\partial V}{\partial r}=\overline{\left(\sigma_x\frac{\partial U}{\partial x}\mathbf{i}+\sigma_y\frac{\partial U}{\partial y}\mathbf{j}+\sigma_z\frac{\partial U}{\partial z}\mathbf{k},\mathbf{n}\right)} \quad (r=a) \qquad \ldots\ldots(29)$$

in place of (18), we find that the formulae for the field are exactly the same as with the dielectric case except that ϵ_x, ϵ_y, ϵ_z, ϵ_0 and ν are replaced by σ_x, σ_y, σ_z, σ_0 and ν', respectively.

A subtle point arises in regard to the calculation of the couple. We use Livens' tensor, as before, and find that *dielectric non-linearity as well as conductive non-linearity* is involved, since this tensor contains ϵ. Taking this into account in equation (24), the couple is found to be

$$G_z'=G_z(\tau_x,\tau_y,\tau_z,\nu')+\frac{27}{10}\epsilon_0(\nu-\nu')F^4a^3\sin 2\gamma\left(\frac{1}{2+\tau_x}-\frac{1}{2+\tau_y}\right)$$
$$\times\left[\cos^2\gamma\left(\frac{3\tau_x^2+4}{(2+\tau_x)^2}\right)+\sin^2\gamma\left(\frac{3\tau_y^2+4}{(2+\tau_y)^2}\right)\right], \qquad \ldots\ldots(30)$$

where $G_z(\tau_x, \ldots \nu')$ is (25) with τ_x, τ_y, τ_z and ν' replacing K_x, K_y, K_z and ν.

We may mention incidentally other features of the steady-current problem that are of interest. The quantitative difference between the two types of non-linearity that must exist in general implies the existence of space charge, a formula for which in terms of $\nu-\nu'$ and $\mathbf{E}_0$ is readily obtainable by consideration of the equation of continuity and Poisson's equation. In the limit $\tau_x\to\tau_y\to\tau_z\to\infty$, the problem degenerates to the non-linear analogue of that of an insulated conducting sphere, or conducting plate with a hemispherical boss, influenced by a uniform field (Jeans 1943). There is, of course, no couple in this case.

§ 5. DISCUSSION

We shall focus our attention upon the solution of the main problem. Similar comments will apply to the solutions of the analogous problems.

The most significant feature of the couple (25) is that, as one would expect, the amendment due to the non-linearity is proportional to F^4, so that there is a relatively larger effect the larger the applied field. The fact that only even powers of F occur shows that the results hold for an alternating applied field provided the frequency is not so high as to require the problem to be treated on an electromagnetic basis. A rough criterion is that the wavelength should be much greater than a. The couple is proportional to the volume of the sphere as in the linear case, a fact which is less obvious *a priori* since, other things being equal, one might have anticipated a relatively larger non-linear effect for a smaller sphere, the gradient of E_0 close to the sphere being greater in that case. Finally, we observe that the dependence upon γ is more complicated than in the linear case. It is evident that the maximum value of the couple is no longer when $\gamma=\pi/4$, although the deviation is small.

A dielectric may obey equation (1) through *intrinsic non-linearity* or through *electrostriction*. An intrinsically non-linear dielectric is one in which ϵ is an explicit function of E. It is well known that polar liquids have small but appreciable non-linearity of this type. In the case of electrostriction, the density is a function of E, and ϵ is a function of E because it is a function of density.

In the case of intrinsic non-linearity equation (1) is obtained in a straightforward manner. It is assumed, either fundamentally or on the basis of molecular theory, that it is a good approximation to the correct relationship between ϵ and E. To establish (1) for an electrostrictive dielectric is less simple. A new and rigorous theory of fluid electrostriction has been given by Brown (1951), and the fundamental result, holding when the only external forces are electrostatic, is

$$\int_{\omega_1}^{\omega_2} \frac{dp(\omega)}{\omega} = \frac{1}{8\pi}\left\{E_2^{\,2}\left(\frac{\partial\epsilon(\omega)}{\partial\omega}\right)_2 - E_1^{\,2}\left(\frac{\partial\epsilon(\omega)}{\partial\omega}\right)_1\right\}, \qquad \ldots\ldots(31)$$

where ω is the density, p the pressure that would exist with the same distribution of ω but no field, and the suffixes 1 and 2 signify any two points in the electrostatic field. The evaluation of ω requires a knowledge of the function of state of the fluid, which varies from case to case, but since we are assuming that the electrostriction is small, all cases may be handled in terms of the bulk modulus k, defined by

$$dp = k(d\omega/\omega), \qquad \ldots\ldots(32)$$

although, of course, k itself depends upon the nature of the change of state associated with the increment $d\omega$.

We shall suppose that 1 is a place where $E=0$ and 2 any other place, and shall drop the suffix 2. Working to the first power of $\omega-\omega_1$, we find from equations (31) and (32) that

$$\omega = \omega_1 + \frac{E^2\omega_1^{\,2}}{8\pi k}\frac{\partial\epsilon}{\partial\omega}. \qquad \ldots\ldots(33)$$

If ω_0 is the density with no field, ω_1 is determined in terms of ω_0 by the condition of mass conservation,

$$\int_v \omega\, dv = \int_v \omega_0\, dv, \qquad \ldots\ldots(34)$$

where v is the volume of fluid enclosed. For an open system ($v=\infty$), we easily find, since for large distances R, $E=\mathrm{O}(R^{-2})$, that equation (34) leads to

$$\omega_1 = \omega_0. \qquad \ldots\ldots(35)$$

The change in ϵ due to the change of density is, to the first power of $\omega-\omega_0$,

$$\epsilon-\epsilon_0 = \frac{\partial\epsilon}{\partial\omega}(\omega-\omega_0), \qquad \ldots\ldots(36)$$

and with the use of equations (33) and (35), this becomes

$$\epsilon=\epsilon_0\left(1+\frac{1}{8\pi}\frac{\omega_0^2}{k\epsilon_0}\left(\frac{\partial\epsilon}{\partial\omega}\right)^2E^2\right), \qquad \ldots\ldots(37)$$

which is the same as equation (1), and gives the value of ν. In particular cases, $\partial\epsilon/\partial\omega$ may be calculated from the Mosotti equation,

$$\frac{\epsilon-1}{\epsilon+2}=A\omega, \qquad \ldots\ldots(38)$$

where A is a constant.

Specimen calculations have been made for polar liquids, using values of ν obtained experimentally by Malsch (1928), and for non-polar fluids subject to electrostriction but assumed to be otherwise linear, using values of A quoted by Fröhlich (1949). It is found that the intrinsic non-linearity of polar liquids modifies the couple by amounts larger than electrostriction by factors of the order 10^6 or 10^7. The modifications are, in fact, of the same order as the modifications of electrostatic capacity in Malsch's experiments, and are probably measurable. It seems most unlikely, however, that the effect of electrostriction could ever be detected, much less confused with that of intrinsic non-linearity.

We shall now examine our initial assumptions that the solid dielectric is perfectly linear and rigid. If it was non-linear and rigid, we could apply a perturbation process to the equation for the potential (which reduces to equation (16) in the linear case) and spatial derivations of the field strength would occur in the perturbation term. But these would be zero since the zero-order field is uniform, and it follows that the first-order perturbation would be zero. If the solid was non-rigid, there might be a first-order effect, as the shape of the dielectric boundary would be altered through solid electrostriction, with implications both to the potential problem and to the calculation of the mechanical action. It seems safe to assume, however, that the effect of solid electrostriction would be of similar order to the effect of fluid electrostriction, and this has already been found to be negligible. Our general inference, therefore, is that there is no likelihood of the limitations imposed by the assumptions affecting the practical significance of our results, provided non-linearity of the solid is not so large that the second-order perturbation of the internal field is comparable with the first-order perturbation of the external field.

On the basis of the foregoing discussion it seems reasonable to suggest that our results could be used as the basis of an experimental method for determining ν for polar liquids. In his condenser method, Malsch used fields of 250 kv cm^{-1} and found modifications of capacity of the order of 1%. As already indicated, there would be a modification of couple of the same order if such a field was used in the method we propose. Such a modification should certainly be measurable, and the relative value of our method would probably depend on the degree of accuracy that could be attained.

There is, however, a theoretical difficulty. Our calculation of the field and couple assumes that there is no double layer. In practice a double layer would almost certainly be present, and its strength would probably depend upon that of the applied field. It is therefore to be expected that a non-linear effect would occur from this cause, and there is nothing upon which to judge its magnitude. The same difficulty, however, is inherent in Malsch's method, for the presence of a double layer will also give non-linearity of capacity. We may in fact note that agreement between Malsch's results and the predictions of molecular theory (O'Dwyer 1951) is hardly better than of order of magnitude. Molecular theory, of course, has its own difficulties, and it is not easy to locate the source of discrepancies. In view, however, of the difficulties that face the subject generally, a new experimental method such as we suggest should be welcome notwithstanding its drawbacks.

In the case of the analogous paramagnetic experiment, the difficulty of the double layer does not occur. With the steady-current experiment, however, there would be the additional difficulty of Joule heating, and convection would probably render the experiment useless.

ACKNOWLEDGMENT

The author is indebted to Dr. R. Fürth, of Birkbeck College, University of London, for valuable criticisms of an earlier version of this paper.

REFERENCES

BROWN, W. F., JR., 1951, *Amer. J. Phys.*, **19,** 290.
CADE, R., 1953, *Proc. Phys. Soc.* B, **66,** 557; 1954, *Ibid.*, **67,** 689.
FRÖHLICH, H., 1949, *Theory of Dielectrics* (Oxford : University Press).
JEANS, J. H., 1943, *Electricity and Magnetism* (Cambridge : University Press).
LIVENS, G. H., 1926, *Theory of Electricity* (Cambridge : University Press).
MALSCH, J., 1928, *Phys. Z.*, **29,** 770.
O'DWYER, J. J., 1951, *Proc. Phys. Soc.* A, **64,** 1125.
TRANTER, C. J., 1950, *Quart. J. Math.* (2) **1,** 1.

Modulation of the Surface Conductance of Germanium and Silicon by External Electric Fields

By G. G. E. LOW
Department of Physics, University of Reading

Communicated by R. W. Ditchburn; *MS. received 7th September* 1954
and in amended form 25th October 1954

Abstract. A method is described which enables the modulation of surface conductance by capacitively applied electric fields to be investigated on single crystal specimens. Experimental results are given for n-type and p-type germanium and for p-type silicon. The observed conductance changes and their time dependence provide information concerning the surface barrier and the relaxation phenomena associated with departures from electronic and ionic equilibrium.

---

§ 1. Introduction

The effect of external electric fields on the conductance of thin films of germanium and other semiconducting materials has been examined by Shockley and Pearson (1948). In their experiments the electric field was applied between the film and a metal electrode which together formed a parallel plate condenser (figure 1 (*a*)). In this way charge carriers could be induced in the semiconductor and these were expected to take part in longitudinal conduction. The experiments showed, however, that the change in conductance of the film was much less than that which would be expected if all the induced charge possessed a mobility equal to that of carriers in the bulk of the film. This discrepancy was interpreted as being a consequence of surface states which render most of the induced charge immobile. Experiments of this kind can thus yield information concerning the surface states. The films examined by Shockley and Pearson were produced by evaporation in vacuum. Large single crystals of germanium and silicon are now available, and in view of the general interest in their surface properties, it is desirable to perform similar surface conductance experiments on such crystals. In the present paper a method is described which enables this to be done without the preparation of very thin specimens. Measurements have been carried out on germanium and silicon.

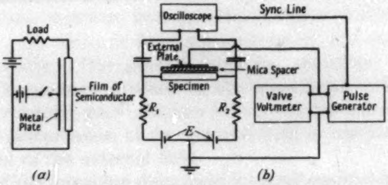


Figure 1. (*a*) Shockley and Pearson's method for measuring the modulation of surface conductance by capacitively applied electric fields (Shockley 1950), and (*b*) the method described in the present paper.

§ 2. EXPERIMENTAL METHOD

For specimens thick enough to be handled conveniently, only small fractional changes in conductance occur even when high external fields are applied. The difficulty of measuring such small changes is increased by drifts of specimen conductance arising from temperature variations. The method used during the present experiments is shown diagrammatically in figure 1 (*b*). Square voltage pulses are applied between the external plate and earth. Initially, in the absence of a sweeping current, R_1 and R_2 are adjusted until no signal appears on the oscilloscope. This corresponds to balancing a bridge circuit of which the two halves of the specimen form two arms. In the presence of a sweeping current, a standing e.m.f. exists between the ends of the specimen. This e.m.f. is, of course, a function of the conductance of the specimen but the response of the oscilloscope amplifiers is such that slow drifts in conductance are not recorded. However, when a pulse is applied to the external plate the conductance of the specimen and consequently the e.m.f. across it changes rapidly. This change δV is recorded by the oscilloscope as a function of time. The corresponding change of specimen conductance is then given by

$$\delta G = -\frac{R+R_1+R_2}{IR^2(R_1+R_2)}\,\delta V$$

where R is the resistance of the specimen and I is the sweeping current. The capacitance between plate and specimen is known and the induced charge can thus be calculated. As the time constant for the charging up of this system is much less than 1 μsec, the transients occupy only a negligible fraction of the 250 μsec period observed on the oscilloscope. In view of this, the balance settings of R_1 and R_2 are not critical.

For the experiments described, R_1 and R_2 were about 1000 ohms and E was adjusted to give a sweeping current of about 5 mA. The specimens were in the form of filaments approximately 2 cm long and 0·05 × 0·2 cm cross section, their resistance being about 2000 ohms. The soldered end contacts had negligible resistances and were tested for the absence of carrier injection by means of a Many bridge (Many 1954). A mica insulating spacer (about 10 μ thick) was placed between the external plate and the specimen. The use of such a spacer made it possible to induce much greater surface charges than could be obtained with air as an insulator. Experiments without the spacer were also carried out and surface charges up to the limit imposed by the dielectric strength of air were induced. These measurements showed that the conductance changes described below are not associated with the mica itself. The external plate did not extend to the ends of the filament, as it was desired to avoid effects due to less thoroughly etched areas of the specimen near the soldered contacts. Because of this a correction involving the ratio of the length of the specimen to the length of the plate had to be applied. The measurements were carried out in darkness and the ambient atmosphere was dried with P_2O_5.

§ 3. RESULTS

The results obtained for etched surfaces on n-type and p-type germanium will be described first. The application of a voltage pulse to the external plate results in an almost instantaneous change of conductance of the specimen. During the next 100 to 200 microseconds this change decays to a quasi-stable value.

The decay is approximately exponential with a time constant, on the specimens examined, of between 10 and 50 μsec. It was observed by another method (described below) that this quasi-stable value eventually decays to zero, or nearly zero, with a time constant of the order of seconds. Typical results are shown in figure 2. The ordinates represent conductance changes in reciprocal ohms relating to a square of surface. The circles refer to the initial conductance changes and the dots to the quasi-stable values. The corresponding induced charge densities are plotted as abscissae. Three n-type and two p-type specimens with resistivities from 7 to 15 ohm cm were examined and gave results similar to those shown. It can be seen that n-type and p-type germanium do not behave in a complementary manner. Figure 3 shows details of the conductance changes observed near the origin on one of the n-type specimens. All measurements were carried out within one hour of treating the surface under examination with CP4 etch.

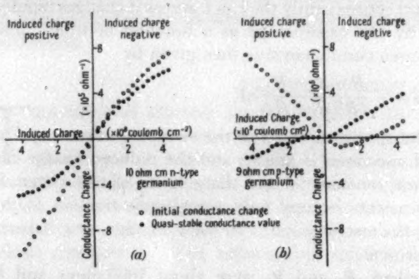


Figure 2. Conductance changes observed for etched surfaces on (*a*) n-type germanium, and (*b*) p-type germanium.

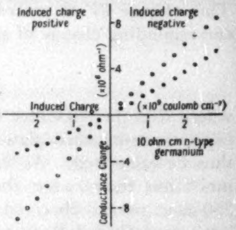


Figure 3. Conductance changes observed near the origin on an etched n-type germanium specimen.

The effect of temperature on the initial conductance changes and on the quasi-stable values was examined for an n-type germanium specimen. Both these quantities decrease with increasing temperature which indicates that surface conduction by ions does not play any important part. The interpretations below are therefore based on electronic models.

Figure 4 shows results obtained for sand-blasted surfaces on n-type and p-type germanium. For such surfaces no decay from an initial conductance change to a quasi-stable value is observed, i.e. the conductance change remains approximately constant throughout the application of a pulse. Similar behaviour is found for low lifetime silicon surfaces (7 ohm cm, p-type material). The results obtained for these surfaces are plotted in figure 5. The changes of conductance on these low lifetime surfaces show slow decays to zero, or nearly zero, similar to those found in the case of etched germanium surfaces.

As the slow decays of conductance take place with time constants of the order of seconds, they cannot be investigated by the method described above. An ordinary Wheatstone bridge was used and, as drifts of specimen conductance due to temperature variations affected the balance of the circuit, the results obtained by this method were of a lower accuracy. These slow decays were found to depend on the nature of the surrounding atmosphere. In the case of one etched n-type germanium specimen, for example, removal of the drying agent from the specimen container increased the rate of the decay by a factor of about five, whilst changing the ambient of the specimen from dry air to a rough vacuum decreased the rate by approximately the same factor.

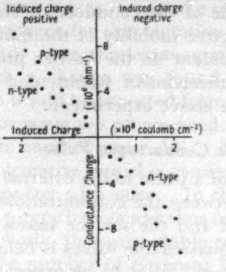


Figure 4. Experimental results for sand-blasted germanium surfaces.

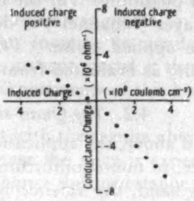


Figure 5. Experimental results for low life-time silicon surfaces.

§ 4. Discussion

The observations described above may be interpreted in the following manner:

4.1. *The Initial Conductance Changes Observed on Etched Germanium Surfaces*

The application of a voltage pulse to the external plate gives rise to an electric field which is propagated through the semiconducting specimen. This results in a charge transfer towards the surface mainly by means of majority carriers since, in the specimens used, the minority carriers are of almost negligible importance in bulk conduction. This conduction process is governed by the time constant involved in charging up the capacitance of the system. Thus, after a period of the order of $0{\cdot}1\,\mu$sec the bulk of the semiconductor is screened by charge near its surface, and no electric field exists in the interior. Near the semiconductor surface, however, the concentrations of electrons in the conduction band, the valence band and the surface states are not in mutual equilibrium.

The change in the number of majority carriers near the surface, resulting from the charge transfer described above, gives rise to an initial change of conductance. In the case of n-type germanium the initial changes observed (figure 2 (*a*)) are consistent with the above arguments in that their directions correspond to an induced charge provided by electrons. Similarly, except for large negative induced charges, p-type germanium shows initial conductance changes which correspond to an induced charge provided by holes (figure 2 (*b*)). The discrepancy observed for large negative induced charges will be discussed in the following section. The above arguments do not depend on any specific assumptions concerning the potential energy profile near the surface.

On the basis of this mechanism it is to be expected that the ratio of the initial conductance change to the induced charge should yield a value for majority carrier mobility. This is in approximate agreement with experiment. For small induced charges on n-type germanium the ratio observed is approximately 3500 cm^2 volt^{-1} sec^{-1} (figure 3), while in the case of p-type germanium a value of roughly 1900 cm^2 volt^{-1} sec^{-1} is found (figure 2(*b*)). Even apart from experimental errors, complete agreement with the accepted mobility values is not expected since the mobility of charge carriers near a surface is probably less than that of carriers in the bulk. Furthermore, a breakdown of the assumption that the entire initial charge transfer is by means of majority carrier conduction would also lead to a decrease in the magnitude of the conductance changes observed. There is also a possibility that the effective mobility of the minority carriers in the inversion layer ('channel') is dependent on the barrier profile and is thus affected by the applied pulse. This mechanism would tend to increase the apparent mobility as evaluated from the above experiment.

4.2. *The Quasi-stable Conductance Values*

As described above, the application of a pulse to the external plate gives rise, near the surface, to non-equilibrium between the concentrations of electrons in the conduction band, the valence band and the surface states. The decay of the initial conductance changes to the quasi-stable values is believed to be associated with the restoration of equilibrium. It is to be expected that the time constant of these decays should be dependent on the recombination velocity at the surface. This was confirmed by a series of experiments during which the recombination velocity of a specimen was varied by different chemical surface treatments.

A quantitative consideration of the quasi-stable conductance values observed on n-type germanium indicates that the potential energy profile at a germanium surface in contact with mica differs from that which is encountered at a free surface. This can best be understood by first postulating that a surface barrier of the normal type exists and then showing that such a barrier fails to account for the experimental results. Positive charge induced on an n-type specimen would lead to a higher and thicker barrier. In principle, the consequent conductance change could be either a decrease or an increase depending on the relative importance of (*a*) the decrease in the number of conduction electrons due to the thicker barrier, and (*b*) the increase in the number of holes in the inversion layer (if such a layer exists). Experiment shows that for induced positive charge on n-type germanium the quasi-stable conductance change is a decrease (figure 2(*a*)). It is concluded, therefore, that the loss of electrons predominates. However, for large induced charges, this decrease is greater than can be understood on the basis of the normal barrier postulated above. By inducing positive charge the barrier height may be raised from the assumed normal position (figure 6(*a*)) to a maximum limited by the close approach of the valence band at the surface to the Fermi level (figure 6(*b*)). The corresponding decrease in conductance can be calculated. Neglecting the change in the number of holes in the inversion layer (which would increase the conductance) and assuming an initial

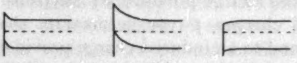


Figure 6. (*a*) Normal barrier assumed to exist at the free surface of n-type germanium, (*b*) position of maximum barrier height, and (*c*) suggested potential energy profile at the surface of n-type germanium when in contact with mica.

barrier height of only 0·2 ev and an electron mobility of 3500 $\mathrm{cm}^2\ \mathrm{volt}^{-1}\ \mathrm{sec}^{-1}$, the calculated maximum conductance decrease is some seven times less than the largest of the quasi-stable values observed experimentally. This difficulty may be resolved by assuming that the surface, when in contact with mica, is initially in the condition shown in figure 6 (*c*), i.e. that the concentration of conduction electrons is greater near the surface than in the bulk of the specimen.† Some further evidence for this tentative assumption was obtained by testing for a change of steady state conductance when mica was placed in contact with a specimen. Such a change was found. An increase in conductance was in fact observed, as would be expected from the present model, if the inversion layer at the free surface is not too pronounced. Experiments of this kind will be more fully discussed in another paper.

The present tentative model is consistent with the results obtained for p-type germanium in contact with mica. In this case the mica is expected to *increase* the height of the p-type barrier and thus to enhance the importance of the inversion layer. Consequently, the quasi-stable conductance values should be controlled by changes in the number of electrons within this layer, in agreement with the results shown in figure 2 (*b*).

The anomalous behaviour of the initial conductance changes observed on p-type germanium for large negative induced charges calls for further discussion. Such induced charges correspond to increases in p-type barrier height. It is possible that, for large barrier distortions, some form of break-down through the barrier occurs almost at once and a partial return to equilibrium is brought about very rapidly. Thus, the initial conductance changes observed on the oscilloscope would not differ much from the quasi-stable values. The absence of a similar effect for large positive induced charges on n-type germanium may be attributed to the different initial condition of these surfaces when in contact with mica (figure 6 (*c*)). In accordance with the present model the n-type barrier would suffer a much smaller distortion than the p-type barrier.

4.3. *Silicon and Sand-blasted Germanium Surfaces*

The conductance changes observed on low lifetime silicon surfaces (figure 5) and sand-blasted germanium surfaces (figure 4) have a common feature in that they remain practically constant throughout the duration of the pulse. This is to be expected as the low lifetimes of minority carriers near the surfaces of these specimens would prevent significant departure from equilibrium in the distribution of electrons in the conduction band, the valence band and the surface states. The changes of conductance of both sand-blasted n-type and sand-blasted p-type germanium are alike in sign and magnitude; the sign corresponding to an induced charge provided by holes. This suggests that the properties of these surfaces are no longer governed by the impurity content of the bulk material but are primarily determined by the disorder arising from the sand-blasting. The direction of the conductance changes observed on p-type silicon indicates that adjustments in the number of holes near the surface predominate, i.e. that there is no pronounced inversion layer in this case.

---

† The author is indebted to Professor J. Bardeen for drawing his attention to this possibility.

4.4. *The Slow Decay of the Quasi-stable Conductance Values*

A tentative explanation of the decay of the quasi-stable conductance values to zero is as follows. The conductance measurements described above imply that the potential energy profile at the surface of germanium is disturbed by the application of an electric field. A change in this profile will vary the effective work function of the surface. Brattain and Bardeen (1953) found that the work function is dependent upon the surrounding gas atmosphere. Consequently, any change in the work function should result in a disturbance of the equilibrium between the surface and the gas. It is suggested that, after the application of an external field, this equilibrium is restored by a change in the number of ions on the surface. If the applied field is in such a direction as to make the surface positive an increase in the number of negative ions (or a loss of positive ions) would result. In consequence, the semiconductor would be screened from the disturbing field and the surface barrier would return to its original condition. Hence, the conductance changes should decay to zero with a relaxation time dependent on the rate of loss or gain of surface ions. The observation that the slow decay process is sensitive to the ambient atmosphere provides some support for this.

§ 5. SUMMARY AND CONCLUSIONS

The sudden application of an external field to an etched germanium surface produces an almost instantaneous change of surface conductance. For small fields these changes are in a direction which corresponds to an induced charge provided by majority carriers. The initial conductance changes decay to quasi-stable values with time constants dependent on the surface recombination velocities of the specimens, and further to zero with time constants of the order of seconds. The first of these relaxation phenomena results from the re-establishment of electronic equilibrium at the surface. The second is due to ionic processes which remain to be investigated in detail. From the magnitude of the quasi-stable conductance values observed it is possible to infer that the potential energy profile at the surface of n-type germanium, when in contact with mica, differs materially from the configuration encountered at a free surface. The application of a pulse to low lifetime silicon surfaces or to sand-blasted germanium surfaces produces changes of conductance which remain approximately constant throughout the duration of the pulse. The absence of the fast decay is interpreted as being a consequence of the low surface lifetimes.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for placing research facilities at his disposal, Drs. H. K. Henisch and P. C. Banbury for encouragement and helpful advice, Professor J. Bardeen for a valuable discussion, and the New Zealand Defence Research Organisation for financial support.

REFERENCES

BRATTAIN, W. H., and BARDEEN, J., 1953, *Bell Syst. Tech. J.*, **32**, 1.
MANY, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).
SHOCKLEY, W., and PEARSON, G. L., 1948, *Phys. Rev.*, **74**, 232.

Measurements of Injection Ratio of Point Contacts on Germanium

By P. C. BANBURY and J. HOUGHTON
Department of Physics, University of Reading

Communicated by R. W. Ditchburn; MS. received 22nd June 1954
and in amended form 24th September 1954

Abstract. Measurements are reported of injection ratio γ on various specimens of n-type germanium. γ was found to be insensitive to the nature of the contact material, to the contact thrust, and to the carrier concentration of the germanium over the limited range investigated. In all cases measured, γ decreased with increasing emitter current over the range 0·5 to 30 mA. A slight decrease of γ with increasing humidity of the ambient air was also observed.

§ 1. Introduction

The forward current across a rectifying contact on the surface of n-type germanium may be carried by both holes and electrons. The injection ratio γ is defined as the ratio of the hole current to the total current across the contact. Measurements of γ have been reported by Shockley, Pearson and Haynes (1949), who obtained a value of unity. The evaluation of γ involves the ratio b of the mobilities of electrons and holes, for which these authors took the value 1·5. The value now accepted for germanium at room temperature is 2·1, which reduces the injection ratio observed by them from unity to 0·8. Hogarth (1953) has reported values of γ from about 0·65 to unity. Further measurements on n-type germanium will now be described.

§ 2. Experimental Method

The method employed was that of Shockley, Pearson and Haynes in which the emitting contact is applied to the side of a filamentary specimen. By the application of a longitudinal electric field, the holes injected from this contact are swept along the filament into a region where the change in local conductivity due to their presence is measured. To make allowance for decay the transit time is varied by altering the sweeping current and an extrapolation is made to zero transit time.

In the analysis for the evaluation of γ, it is assumed that longitudinal diffusion currents may be neglected. Therefore the emitter current I_e must be small compared with the sweeping current I_s. The further assumption is made that the injected carriers are uniformly distributed over the cross section of the filament. This would appear to require that the transit time should be large compared with the time taken for the carriers to move across the filament, largely by diffusion, thus imposing an upper limit on I_s in given circumstances. In a real filament, however, the cross sectional distribution of added holes will never be uniform owing to surface recombination. Nevertheless, provided that $I_e \ll I_s$, the measured change in conductance should to a first approximation be independent of this transverse distribution. It was verified from the experimental results that no correlation could be detected between observed values of γ and the size or shape of the specimens used.

The determination of γ requires the measurement of local conductivity and of transit time. The former involves the measurement of the voltage between two potential probes. To eliminate the error arising here due to the presence of floating potentials at the contacts, the probes were soldered to the germanium by a condenser discharge method, as described by Mitchell (1954). The measurement of the transit time requires a knowledge of the voltage drop in the filamen between the emitter and the conductivity probes. A direct measurement of this voltage will be in error by an amount equal to the voltage across the injecting barrier. However, in the condition $I_e \ll I_s$ the transit time may be taken to be inversely proportional to I_s. In practice, an exponential decay law was observed and extrapolation to zero transit time could be made with accuracy, as is seen from the example given in figure 1.

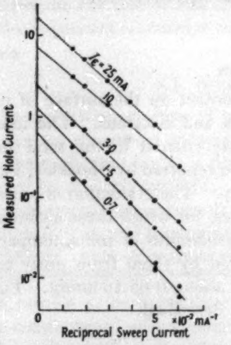


Figure 1. Measurements showing extrapolations for determination of γ (taken from observations of γ with I_e).

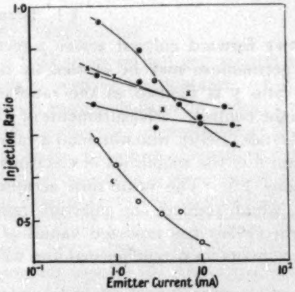


Figure 2. The variation of injection ratio with emitter current (tungsten contact on n-type germanium).

End contacts were applied to the filament by direct soldering and covered with a protective film of polystyrene. The crystal was then etched in CP4 immediately before the application of the point contacts. The assembly was placed in a dark container to eliminate photo-effects. This arrangement also enabled the humidity of the ambient air to be varied. Generally, the experiment was performed using d.c. supplies. For one part of the investigation, however, involving an extended range of emitter currents, it was necessary to use pulse methods, which will be discussed in the appropriate section.

§ 3. RESULTS AND DISCUSSION

3.1. *General Remarks*

In the course of these experiments many observations of γ have been made, and values have been measured ranging from 0·1 to 1·0. By precautions in preparation on any one filament the scatter can be reduced to a much smaller range.

With these precautions, systematically different results were obtained for different surfaces which had been given apparently the same treatment. These differences did not appear to be related to the normal carrier concentration of the specimens which varied from 1×10^{14} cm^{-3} to 7×10^{14} cm^{-3}, nor to the crystal orientation of the face on which the contact was made. High and low values of γ could not be systematically associated with 'clean' and 'dirty' surfaces respectively; for example in one test a surface giving values of γ of the order of 0·75 immediately after mounting, gave values of the order of 0·98 after exposure to laboratory air for several weeks. No unique value of γ for n-type germanium can be quoted, though the value obtained in these experiments normally lay between 0·7 and 1·0.

3.2. *Variation of γ with Contact Thrust*

Since the normal method of applying the emitter gave no accurate control of contact thrust, it was necessary to carry out an experiment to determine the variation in γ over the range of thrusts likely to be applied. For a given contact no change in γ was observed for contact thrusts up to 10 grammes. By the continuous observation of the voltage ΔV between the conductivity probes for a single value of sweeping current I_s it would have been possible to detect a change of 1% in γ.

3.3. *Variation of γ with Contact Material*

The electrical properties of a metal–germanium contact are known to be generally insensitive to the work function of the metal. This can be accounted for by the postulated existence of surface states of such a density that the dependence of barrier height on the metal work function is much reduced. For any finite density of surface states, however, some dependence of barrier height on the metal must remain. Now it is to be expected on the basis of the barrier model that for a given n-type germanium specimen the hole current will be approximately independent of the equilibrium barrier height while the electron current should be exponentially dependent on it. Thus, in the case of contacts for which γ is appreciably different from 0 or 1, γ would be expected to be particularly sensitive to the barrier height and so possibly to changes of work function of the metal. Measurements were carried out to test for systematic differences for different metal contacts.

In one test a comparison was made of the injection at $I = 0·5$ mA for platinum and freshly pointed tungsten contacts. Measurements were taken on twenty contacts for each metal, made near the same point on a filament. Medians corresponding to $\gamma = 0·96$ for platinum and 0·93 for tungsten were obtained. The difference was less than the estimated standard error for this difference and cannot be regarded as significant. On another specimen determinations of γ were carried out for tungsten, graphite and n-type germanium contacts. The values of γ obtained were of order 0·35, the differences for different materials being again not greater than the differences observed between various contacts of tungsten.

In the particular case of a Ge–Ge emitting contact, it is possible to alter the energy distribution of electrons in the emitter continuously without introducing changes of γ associated with a change of contact position. This may be done by using a subsidiary process of injection into the emitter near the contact. A test was carried out using as emitter a well-etched specimen of material with a diffusion range of the order of 0·5 mm which was greater than the spacing between the secondary emitter and the Ge–Ge contact. Using an emitter current of 1 mA and a subsidiary injected current of 3 mA no change in γ ($\sim$0·6) could be observed, heating effects having been eliminated. This negative result indicates that the Ge–Ge contact in question had no detectable transmission to added carriers.

3.4. *Variation of γ with Forward Emitter Current*

Measurements were made to determine the variation of injection ratio with total current. It is a requirement of the method that the emitter current must be small compared with the sweep current, as mentioned in § 2. Thus, in the case of emitter currents which are not small, large sweeping currents must be used. To eliminate heating effects the duty cycle was reduced by suitably timed mechanical switching of the current supplies and the potentiometer circuit. (Repetition frequency 1 c/s; on–off ratio 1 : 10; galvanometer switching delayed and advanced 1/100 sec from beginning and end respectively of the supply pulse.) It was found possible, by the use of this method to extend the range of emitter currents to smaller as well as larger values.

The value of γ was found to decrease with increasing emitter current. Characteristics obtained for five point contacts are shown in figure 2.

This decrease in γ with increasing current may be accounted for in the following way. An increase in emitter current causes an increase in the local hole concentration, and as a consequence of the neutrality condition the electron concentration also increases, thus enhancing the flow of electrons to the metal. Bardeen and Brattain (1949) have pointed out that this gives a condition in which the electron current increases more rapidly than the hole current with applied voltage, and hence γ will decrease as the emitter current is raised. They predicted an upper limit of approximately 10 mA for the current at which this fall of γ should become appreciable. It can be estimated (Banbury 1952) that if the lowering of effective barrier height due to hole space charge in the barrier near the metal is taken into account, the fall in γ with emitter current should in all cases be appreciable for currents of less than 1 mA. It will be seen from figure 2 that under the conditions of measurement here γ is decreasing with increasing current in all cases over the whole of the measured range, down to 0·5 mA.

3.5. *The Variation of γ with Humidity of Ambient Air*

The effect of humidity of the ambient air was observed for contacts on two specimens, one set with good and the other with poor injecting properties. In both cases a decrease in γ was observed as the humidity increased. On the more highly injecting surface ($\gamma \sim 0{\cdot}8$) no systematic change in filament conductance with ambient was observed. Conductance changes ($\sim$5%) accompanied γ changes ($\sim$40%) on the poorly injecting surface ($\gamma \sim 0{\cdot}3$). In the second case, therefore, part of the change in γ with humidity might be accounted for by the flow of the emitter current over the surface of the filament and not through the injecting barrier to the bulk, but this process cannot account for the whole of the change in γ. If a reduction of γ is attributed to a lowering of the barrier height, it might be supposed that the water vapour gives rise to the formation of a surface layer of positive ions resulting in a dipole which tends to lower the barrier.

This would be consistent with observations in this laboratory, on similar specimens, that the presence of water vapour and the application of a positive field normal to the surface are also equivalent in their influence on surface recombination (Henisch, Reynolds and Tipple 1955).

ACKNOWLEDGMENTS

The authors wish to thank Professor R. W. Ditchburn for the provision of research facilities, Dr. H. K. Henisch for advice and encouragement, and the Royal Naval Scientific Service for their support, and for permission to publish this paper.

REFERENCES

BANBURY, P. C., 1952, *Thesis*, University of Reading.
BARDEEN, J., and BRATTAIN, W. H., 1949, *Phys. Rev.*, **75**, 1208.
HENISCH, H. K., REYNOLDS, W. N., and TIPPLE, P. M., 1955, *Physica*, in the press.
HOGARTH, C. A., 1953, *Proc. Phys. Soc.* B, **66**, 845.
MITCHELL, W. H., 1954, *J. Sci. Instrum.*, **31**, 147.
SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R., 1949, *Bell Syst. Tech. J.*, **28**, 344.

The Hall Coefficient, Electrical Conductivity and Magneto-Resistance Effect of Lead Sulphide, Selenide and Telluride

By E. H. PUTLEY

Radar Research Establishment, St. Andrew's Road, Gt. Malvern, Worcs.

MS. received 16*th August* 1954

Abstract. The experimental procedure and results are described for measurements of the Hall coefficient and electrical conductivity of single crystals and natural specimens of PbS, PbSe and PbTe over the temperature range 77–1000°K. Some measurements at 20°K are also described. The magneto-resistance effect was also measured on some specimens at a number of temperatures between 20°K and 300°K.

At high temperatures all specimens show intrinsic behaviour. This will be discussed in detail in a subsequent paper but the behaviour in the extrinsic range, the temperature dependence of the mobility and irreversible effects found at high temperatures are discussed. It is concluded that the general behaviour of the effects considered can be accounted for by the usual semiconductor model.

---

§ 1. Introduction

Measurements of the Hall coefficient and electrical conductivity have been made on a number of specimens of the compounds PbS, PbSe and PbTe over the temperature range 77°–1000°K. Some measurements were also made at 20°K and for some specimens the magneto-resistance effect and the variation of the Hall coefficient with magnetic induction were measured at a number of temperatures between room temperature and 20°K.

In this paper the experimental techniques and the results obtained will be described and some aspects of the results will be discussed. Summaries of some of the conclusions reached for these experiments have already been published (Putley and Arthur 1951, Putley 1952 a, c). Measurements of thermoelectric and thermomagnetic effects have been made on some specimens. These effects are discussed in a separate paper (Putley 1954).

The majority of the measurements were made on single crystals prepared in these laboratories (Lawson 1951, 1952) but some measurements were also made on specimens of natural PbS, galena, and PbSe, clausthalite. There has already been a considerable amount of work published on the properties of these compounds, but until recently few measurements have been made on single crystals. By studying single crystals it has been possible to build up a more complete and consistent picture of the electrical properties of these compounds than previously had been obtained.

§ 2. Experimental Procedure

The crystals were cleaved into rectangular blocks of approximate dimensions 1·0 × 0·25 × 0·1 cm. The edges of these blocks were always parallel to the (100) direction for this is the direction in which the crystals readily cleave. The prepared crystals were mounted in a holder as illustrated in figure 1. Current and potential leads and nickel-chromium–nickel-aluminium thermocouples for measuring the temperature were welded to the crystals using techniques described by Mitchell (1954). The connections made in this way provided low resistance ohmic contacts which were satisfactory over the entire temperature range studied (20° – 1000°K). Platinum wire 0·008 in. in diameter was used for the current leads, while 0·002 in. diameter platinum wire was used for the potential probes and 0·002 in. diameter T_1 and T_2 alloys (similar in their properties to 'Chromel' and 'Alumel' respectively) were used for the thermocouples. In the holder two pairs of potential leads were fitted, one pair for Hall effect and the other for conductivity measurements, and a thermocouple was attached for temperature measurement.

All measurements were made using direct current. All potentials were measured by means of a precision potentiometer of which the smallest subdivision was one microvolt. When a higher sensitivity was required a galvanometer was used to interpolate between the potentiometer settings. By this means potential differences could be measured to a tenth of a microvolt when required.

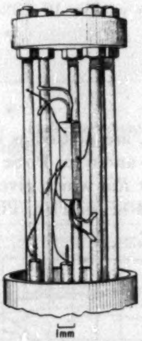


Figure 1 8-pin crystal holder.

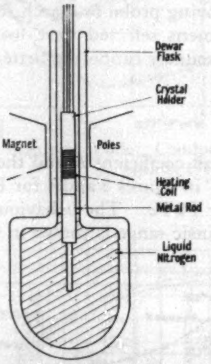


Figure 2. Low temperature apparatus.

For measurements above room temperature the crystal holder was placed inside a furnace made by winding a heater on a silica tube. This was placed between the poles of an electromagnet. A stream of argon was passed over the specimen to reduce irreversible effects, but it was not possible completely to eliminate irreversible changes at very high temperatures. Below room temperature the holder was placed inside a Dewar flask. For studying the magneto-resistance effect or other experiments at fixed temperatures tailed Dewars were used, but for measurement between 77°K and room temperature the arrangement shown in figure 2 was found satisfactory. The temperature of the specimen was reduced by means of a metal rod attached to the holder and dipping into liquid nitrogen and the temperature could be raised to room temperature by means of the heater wound round the specimen holder. Thus, by suitable adjustment, measurements could be made at intermediate temperatures.

The electromagnet enabled a magnetic induction up to about 10 000 gauss to be used. Hall coefficient measurements were usually made with inductions of about 8000 gauss, but the magneto-resistance effect was studied with variable induction. Some measurements were made on the variation of Hall coefficient with magnetic induction.

In making measurements the usual precautions of reversing currents and magnetic fields were observed. This procedure can eliminate all spurious voltages in Hall effect measurement except any associated with the Ettingshausen–Seebeck effect. This source of error was investigated experimentally and found to be negligible (Putley 1952 b). It is considered that the relative accuracy of these measurements should be within 5%. Currents and potential could normally be measured to better than 1% and magnetic induction to about 2%. It is difficult to set a figure on the absolute accuracy because this depends on the homogeneity and the uniformity of the dimensions of the specimen. Attempts were made to select the best available specimens by checking their uniformity using a moving probe to search for any variation of resistance along the length. The specimens selected were usually uniform to within 10–20%, so that the absolute accuracy cannot be better than this.

§ 3. Results

The Hall coefficient R and the conductivity σ, are plotted as log R or log σ against $1/T$ in figures 3 and 4 for PbS, in figures 5 and 6 for PbSe and in figures 7 and 8 for PbTe. The behaviour of the product $R\sigma$, which gives the mobility in the extrinsic range is shown in figures 9, 10 and 11 for PbS, PbSe and PbTe

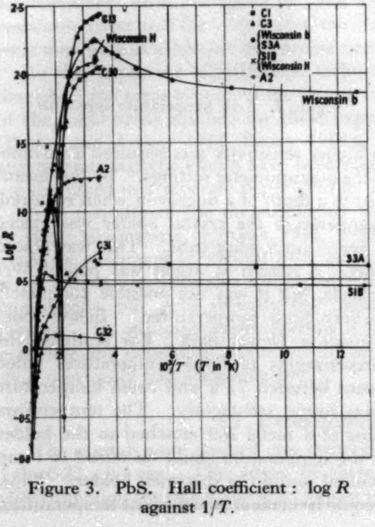


Figure 3. PbS. Hall coefficient : log R against $1/T$.

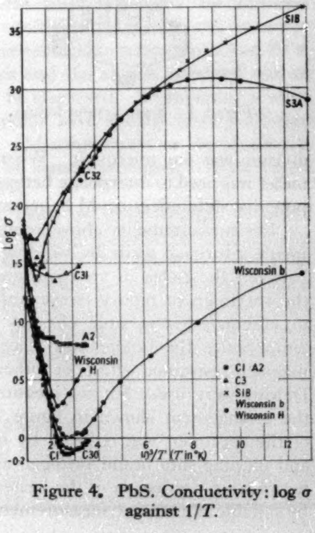


Figure 4. PbS. Conductivity: log σ against $1/T$.

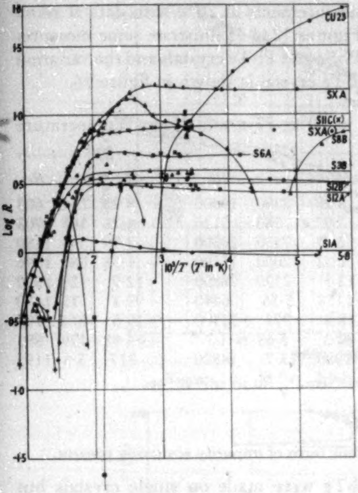


Figure 5. PbSe. Hall coefficient : log R against $1/T$.

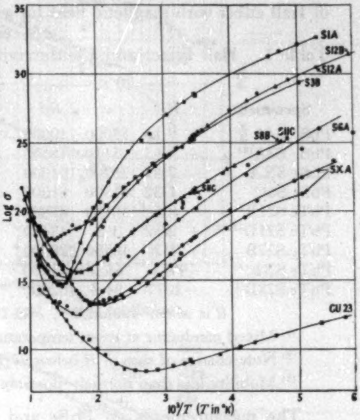


Figure 6. PbSe. Conductivity : log σ against $1/T$.

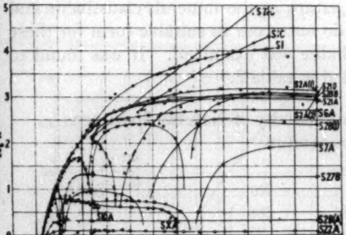


Figure 7. Hall coefficient of PbTe. log R against $1/T$.

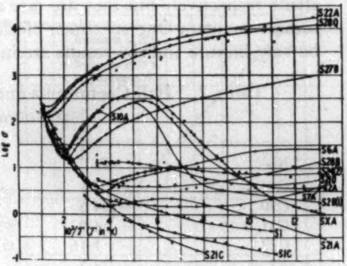


Figure 8. Conductivity of PbTe. log σ against $1/T$.

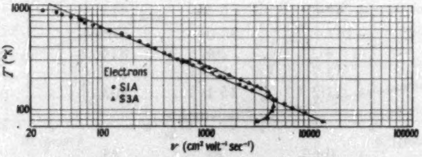


Figure 9. Mobility. PbS single crystals.

respectively. Table 1 gives results of measurements at 20°K with data at room temperature and 77°K for comparison. Figures 12 to 15 illustrate some measurements of the magneto-resistance effect in PbSe and PbTe crystals and the variation of Hall effect with magnetic field for a PbTe crystal is shown in figure 16.

Table 1. Hall Effect and Conductivity at 20°K, 77°K and Room Temperature

	20°K			77°K			290°K		
Specimen	R	σ	$R\sigma$	R	σ	$R\sigma$	R	σ	$R\sigma$
PbS S4A	− 9·16	15000	140000	− 8·28	2390	19000	− 8·25	73·1	603
PbS S3A(3)	+ 3·65	5110	18600	+ 3·97	786	3120	+ 4·26	165	703
PbSe SXA	− 2·87	35200	101000	− 2·87	7150	20500	− 2·99	394	1180
PbSe S6C	+ 1·32	61500	81200	+ 1·31	12050	15790	+ 1·58	884	1396
PbTe S21B(3)	− 8·7	10100	87800	−13·5	2130	28800	−12·7	120	1520
PbTe S21D(1)(3)	+ 2·32	1·5	3480	+1158	5·56	6440	−99·1	15	1490
PbTe S27B	+21·2	10760	228000	+18·8	977	23400	+19·6	40·5	794
PbTe SXE(2)	−17·3	65·3	1130	+ 0·3	5·68	1·7	+ 4·98	179	893
PbTe S28B(1)	−1077	94·4	102000	−1080	13·7	14820	−217	5·5	1195

R is in cm^3 coulomb^{-1}, σ in ohm^{-1} cm^{-1}, $R\sigma$ in cm^2 v^{-1} sec^{-1}.

(1) Mixed conductor at room temperature.

(2) Note change of sign of R below 77°K.

(3) Mobility less than normal—possibly some form of impurity scattering present.

The measurements on PbSe and PbTe were made on single crystals but the majority of the measurements on PbS were made on natural specimens of galena. Many natural specimens were examined before suitable ones were found. The results of measurements on natural specimens are summarized in table 2. PbSe and PbTe also occur in nature as the minerals clausthalite and altaite respectively but they are rare and do not occur in suitable form for these measurements. One specimen of clausthalite was examined. It was found to be very impure and the results are included in table 2.

Table 2. Hall Coefficients and Conductivity of Natural Specimens at Room Temperature

		Galena		
No.	R	σ	$R\sigma$	Source
C1	+ 219	0·94	206	Sardinian galena
C30	+ 94·5	0·96	91	
C31	− 4·92	30·9	152	
C32	− 1·16	202	234	
A2	− 23·5	6·28	148	
Wb	−143	1·13	162	Wisconsin galena
WH	−124·5	4·09	510	
J5	+154	3·57	550	Joplin galena
J4	+ 4·49	59·7	268	
J6	+ 62·8	6·58	413	
Urals	−147	32·0	470	Source in opposite column
Lead Hills	− 4·2	208	874	
Cornwall	− 3·7	164	607	
Saxony	+ 53·2	15·1	803	
		Clausthalite		
Hartz Mountains	+ 0·65	70	45	

R is in cm^3 coulomb^{-1}, σ in ohm^{-1} cm^{-1}, $R\sigma$ in cm^2 v^{-1} sec^{-1}.

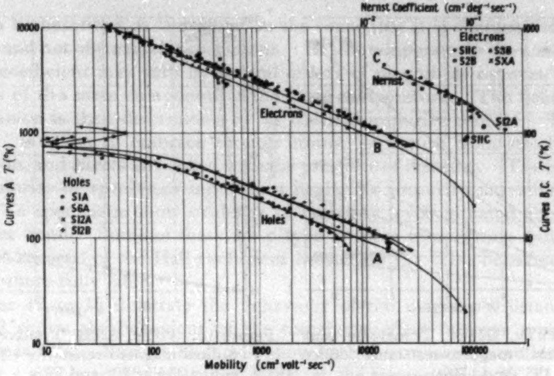


Figure 10. PbSe. Mobility and Nernst coefficient.

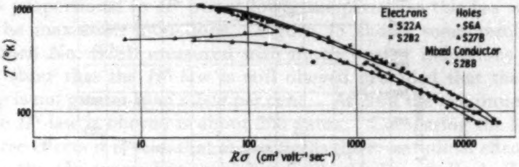
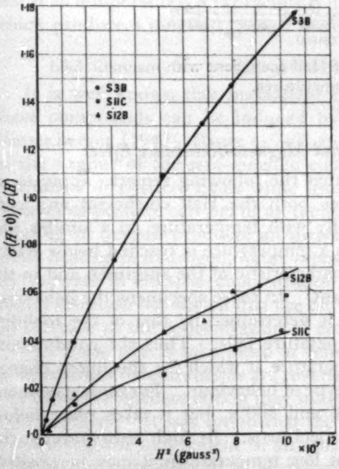


Figure 11. Mobility. PbTe.



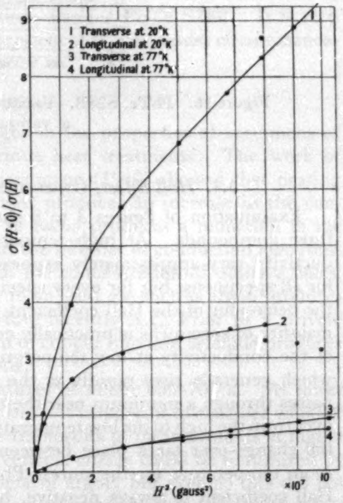


Figure 12. Magneto-resistance effect in PbSe.

Figure 13. PbTe, S28B. Magneto-resistance effect. Transverse and longitudinal at 77° and 20°K

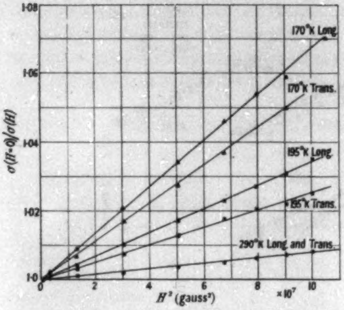


Figure 14. PbTe, S28B. Transverse and longitudinal magneto-resistance effect at 290°, 195° and 170°K.

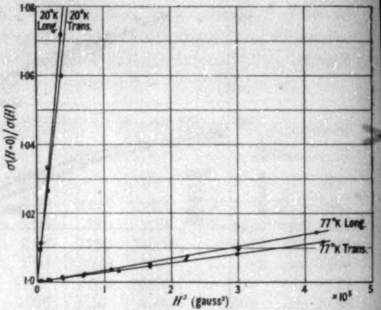


Figure 15. PbTe, S28B. Transverse and longitudinal magneto-resistance effect with weak inductions at 20° and 77°K.

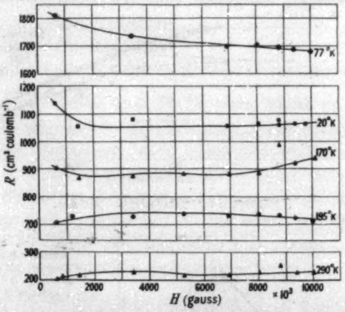


Figure 16. PbTe, S28B. Variation of Hall coefficient with magnetic field at various temperatures.

3.1. *Hall Effect, Conductivity and Magneto-Resistance Effect*

Examination of figures 3 to 8 show that the following remarks apply to all three compounds. At high temperatures both the Hall coefficient and conductivity vary approximately exponentially with temperature, in a similar way for all specimens, but for every specimen a temperature is reached below which the behaviour of the Hall coefficient is characteristic of the specimen and in the majority of specimens is practically constant. In these specimens the behaviour of the conductivity at low temperatures is determined by that of the mobility which generally rises rapidly as the temperature falls. Thus the conductivity passes through a minimum near the temperature at which the specimens change over from the high to the low temperature type of behaviour. For most specimens the change-over takes place between 400 and 500°K, but it takes place below room temperature for the purest PbTe specimens. At high temperatures the Hall coefficient is always negative, but at low temperatures it may be positive or negative. At high temperatures the behaviour of all specimens is so similar that it can be described as 'intrinsic' in the sense that it is characteristic of the substance and not of particular specimens. At low temperatures the conductivity and Hall coefficient may vary by several orders of magnitude between different specimens of the same compound at the same temperature. The behaviour of each specimen is then determined by its residual impurity content. The term 'impurity' is taken to embrace foreign atoms, departure from stoichiometric composition, and effects associated with the presence of disorder. The behaviour of the majority of specimens indicates a negligible impurity activation energy. Some PbTe specimens show evidence of impurity levels situated about 0·1 ev from either band. Some of these are p-type at room temperature and show an anomalous reversal of the Hall coefficient below 150°K. This behaviour will be discussed more fully later.

Figures 12 to 15 illustrate the behaviour of the magneto-resistance effect. Figure 12 shows the behaviour of the transverse effect in some typical PbSe specimens at 77°K. Figures 13 and 14 show both the transverse and longitudinal effects for a specimen of PbTe, No. S28B, at the temperatures 20, 77, 170, 195 and 290°K for inductions up to 10000 gauss. Down to 170°K the change in resistance is proportional to H^2 but at lower temperatures this law is no longer obeyed at the maximum inductions. Figure 15 shows some results, also on PbTe specimen No. S28B measured with much weaker inductions at 20° and 77°K which show that the H^2 law is still obeyed provided that the change in conductivity is not greater than a few per cent. At 20°K the maximum induction at which the H^2 law is obeyed is about 200 gauss. Comparing the longitudinal and transverse effects it is seen that at low fields the longitudinal effect is slightly greater than the transverse but at higher fields the longitudinal effect did not increase so rapidly as the transverse effect. Figure 16 shows the behaviour of the Hall coefficient as a function of induction, also for PbTe S28B. It is seen that at all temperatures R is practically independent of H, even under circumstances which produce a ninefold change in resistance.

3.2. *Irreversible Effects*

It is well known that permanent changes in the properties of specimens of these compounds can be induced by various heat treatments. The work of Hintenberger (1942), Bauer (1940) and Eisenmann (1940) showed that heating in the vapour of the non-metallic constituent produces an increase in the concentration of positive holes while heating *in vacuo* produces a reduction in the number of positive carriers or an increase in the number of conduction electrons. Heating in oxygen produces an increase in the concentration of positive holes, while heating in hydrogen, water vapour or in an inert gas such as argon produces an increase in electron concentration. More recently A.C. Prior in a private communication has shown that the diffusion of copper into lead selenide produces an increase in the concentration of electrons.

All these effects take place at temperatures of about 600–700°K. At lower temperatures measurements are reversible, but it is important to consider whether the existence of these effects may influence the results of measurements at higher temperatures. It was found that the irreversible effects could be minimized by conducting the experiment in an atmosphere of argon, but even so this did not prevent a certain amount of evaporation from the specimen and in experiments taken up to 1000°K up to 10^{18} additional conduction electrons could be produced per cm^3. These additional carriers remain when the crystal is cooled to room temperature and may cause a change of several orders of magnitude in the Hall coefficient and conductivity of the specimen at room temperature before and after the experiment. At first sight it would seem that this would invalidate the results of measurements at high temperatures. When, however, the concentration of new carriers is compared with the concentration of intrinsic carriers at the temperature at which the new carriers are formed it is found that the new carriers do not represent more than 10 to 20% of the total number present, but this does set a limit on the reliability of these results. Scanlon (1953) has suggested that above 500°K the irreversible changes may take place at a sufficiently fast rate to alter the shape of the curves and to render any attempts at deducing intrinsic data from results above 500°K valueless. Our experiments are in agreement with Scanlon's description of the sequence of events which occurs when crystals are taken to progressively higher temperatures but we set the maximum temperature up to which the irreversible changes may be safely neglected at about 800°K and find that under the conditions of our experiments the irreversible effects did not produce changes in the slope of the curves until temperatures greater than 850°K were reached. This effect can be seen in figure 8, specimens S27B and SXA. The time taken to measure our curves was usually five to eight hours, but to check the reliability of these results an experiment was made on a PbTe crystal in which the resistance–temperature curve was measured at a much faster rate using a high speed recording potentiometer. The crystal, mounted in the standard holder, was heated in an argon atmosphere through a number of temperature cycles between room temperature and progressively higher temperatures up to 919°K. The total time for each cycle was 15–20 minutes. At the beginning of the experiment the conductivity had the value indicated on figure 17 at the point A. After each of the first four cycles the conductivity returned exactly to this point. The maximum temperature reached in each cycle is shown

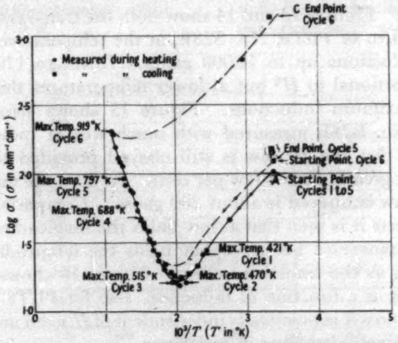


Figure 17. PbTe, S326. Conductivity measured during a series of heating and cooling cycles to progressively higher temperatures. Duration of each complete cycle (room temperature to maximum and back) 15–20 minutes. Cycles 1 to 4 returned to same starting point, cycle 5 ended with σ about 12% greater, cycle 6 produced a 10-fold increase in σ.

on the diagram, the maximum temperature of the fourth cycle being 688°K. The maximum temperature of the fifth cycle was 797°K and at the end of it the conductivity returned to the point B, representing an increase of conductivity of about 12%. A small amount of evaporation appeared to take place during this cycle. The final cycle went up to 919°K. More evaporation was observed and a large irreversible change occurred. The final conductivity (point C) was more than ten times greater than the initial value. Examination of figure 17 shows that for the reversible cycles the points measured on the heating and cooling curves may differ by about 10–15%, but this is attributed to the temperature of the crystal lagging behind that indicated by the thermocouple. It therefore seems unlikely that the measuring time could be reduced further by using pulsing techniques (Ehrenberg and Hirsch 1951).

The cycling experiment shows that perfectly reversible results may be obtained up to 700°K and that appreciable irreversible changes do not take place until temperatures greater than 800°K are reached. Comparison of the curve of figure 17 with those of figure 8 shows that their behaviour in the intrinsic range is comparable. It is therefore concluded that for the bulk of our results the slope of the curves was not influenced by the irreversible changes which took place.

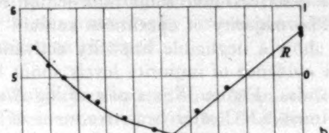


Figure 18. PbTe, S21C. Hall coefficient and conductivity: log R and log σ against $1/T$.

In some cases the irreversible effects may be used to advantage to reduce the concentration of free carriers. Thus by introducing carefully controlled quantities of copper, Prior has produced specimens of PbSe of near intrinsic properties at room temperature. The concentration of carriers in PbTe has been reduced by neutron bombardment. A second process peculiar to PbTe is one of 'annealing'. In this process specimens of PbTe were maintained at a temperature of about 400°K in air for periods varying from one week to three months. The resistance of specimens was measured during this process using a recording potentiometer. It was found that when an impure specimen of PbTe was placed in the oven its resistance rose at first due to the increase in temperature, but after the temperature had become constant the resistance continued to rise until eventually it reached a constant value. On allowing the crystal to cool to room temperature the resistance rose further. This suggested that the effect of the heat treatment was to convert an impure specimen to one of near-intrinsic properties. This suggestion was borne out by Hall effect measurements. This process differs from the better known ones in that it takes place at a much lower temperature and does not appear to depend on the nature of the atmosphere surrounding the crystal. Since it tends to reduce the number of extrinsic carriers it should not affect measurements of the intrinsic properties. Thus in some cases it was found that p-type specimens were converted to n-type despite the oxygen atmosphere. The process appears to be more in the nature of an annealing effect, suggesting the removal of strain or lattice defects.

§ 4. DISCUSSION

4.1. *Intrinsic Properties*

The behaviour in the intrinsic region is still under consideration and will be the subject of a further paper. We have established that the intrinsic energy may be identified with the absorption edge and is thus considerably smaller than the value derived from measurements of Hall or conductivity slopes (Putley 1952 c).

4.2. *Extrinsic Properties*

The measurements described give some information regarding the position of impurity levels. The majority of specimens contain 10^{16} to 10^{18} extrinsic carriers per cm^3 and show a negligible impurity activation energy, but some PbTe specimens show evidence of impurity levels about half way between the full and conduction bands. Figure 18 is a plot of log R and log σ against $1/T$ for PbTe n-type specimen S21C after heat treatment of the type described in § 3.2. Both R and σ give a straight line plot below 150°K and these yield an impurity energy level 0·1 ev below the conduction band. Similar behaviour was found for PbTe S1C (figures 7 and 8).

Further evidence is given by some p-type specimens of PbTe (S2B1, S2A2, S7A, S21A, SXA) which are also shown in figures 7 and 8. It will be seen that between room temperature and about 200°K the number of holes is constant and the conductivity rises as $T^{-5/2}$. On further cooling the concentration of holes falls rapidly, causing the conductivity to pass through a maximum and then to fall. The conductivity curves suggest energy levels about 0·1 ev above the full band but the Hall effect does not behave in a straightforward way. With the reduction in the concentration of holes the Hall coefficient falls to zero and below 150°K becomes negative. Thus in these specimens the Hall coefficient changes sign twice. At high temperatures the normal change of sign is observed on approaching the intrinsic region, but the behaviour at low temperature has not been satisfactorily explained. Similar effects have been observed in the Cu doped PbSe prepared by Prior and in photosensitized PbS and PbTe layers.

4.3. *Mobility*

The log–log plots, figures 9, 10, 11, of $R\sigma$ against T indicate that for n-type specimens at temperatures higher than 100°K $R\sigma$ varies as T^{-n} with $n=2{\cdot}5$ (Putley 1952 a). The order of magnitude in cm^2 $volt^{-1}$ sec^{-1} of the mobilities in these compounds are 500–2000 at room temperature, 10 000–20 000 at 77°K and 100 000 at 20°K. On lowering the temperature below 100°K the mobility continues to increase, but not as rapidly as $T^{-2{\cdot}5}$. Not sufficient data are available to show what law is followed at very low temperatures. The behaviour of some specimens suggests that scattering at lattice defects or impurities may reduce the mobility at low temperatures, but there is no evidence for the Rutherford type of impurity scattering as found in germanium. The results for holes are not as conclusive as those for electrons because the onset of mixed conduction restricts the temperature range over which the properties of holes can be clearly determined. The results for p-type specimens are not inconsistent with a T^{-n} mobility law with n being between 2·5 and 3·0. Devyatkova *et al.* (1952) have measured the mobility in PbSe and found $n=3$ but do not say whether their material was p- or n-type. Petritz *et al.* (1954) have attempted to fit the Mott–Fröhlich mobility formula to galena but find a Debye temperature of about 1100°K necessary to fit the results. A similar value was found when we attempted to fit this formula to the mobility of polycrystalline PbTe (Chasmar and Putley 1951). Since determination of the Debye temperature by measurements of specific heat (Parkinson and Quarrington 1954) and of elastic constants gives values below 200°K it would suggest that the Mott–Fröhlich theory is not applicable to these compounds at the temperatures at which measurements have so far been made. Recently Scanlon (private communication) has shown that using the correct Debye temperature and the scattering formula of Howarth and Sondheimer (1953) and combining with this scattering by acoustical modes a fair approximation to the mobility variation can be obtained.

Table 3. Mobility calculated from Magneto-Resistance Effect compared with Hall Mobility

Compound	Specimen	Temp. (°K)	μ from mag. res.	μ from Hall
PbTe	S32B	290	1075	+ 802
	S28B	20	345000	−112500
		77	31600	− 14250
		170	4590	− 3000
		195	3070	− 2400
		290†	1673	− 1195
	S22A	77	13550	− 20850
PbSe	S12A	300	915	+ 1050
		77	5400	+ 21300
	S12B	77	6500	+ 25000
	S11C	77	5400	− 23000
	S3B	77	12100	− 23400

† Mixed conductor; neither result represents true mobility.

Electron mobility can be calculated from the magneto-resistance effect, for at low field strengths $\delta\sigma/\sigma=CH^2$ where according to classical theory $C=0{\cdot}272\mu^2$, and μ is the mobility. Table 3 shows the value of μ calculated from this formula compared with the value of $R\sigma$. It is seen that the results agree to the order of magnitude but in most cases they differ by a factor of 2–3. The mobility could also be calculated from the Nernst effect since $Q=\frac{1}{2}\boldsymbol{k}R\sigma/\boldsymbol{e}$ (Putley 1954). It follows that in the impurity range Q should vary with temperature in the same way that μ does. This has been checked by plotting values of Q for PbSe on figure 10 where they can be compared with the mobility. It is seen that they both vary with temperature in the same way. Examination of the values found for the mobilities in p-type and n-type specimens (figures 9, 10, 11 and tables

1, 2 and 3) and especially for PbSe, shows that although all specimens follow similar laws the spread in the values between different specimens of the same type is so great that it is not possible to determine the ratio of electron to hole mobilities b by comparing directly the mobilities measured on different crystals. Since p-type specimens always change sign at high temperatures it is clear that the mobility of electrons must be greater than that of holes. It is possible to calculate the mobility ratio from R and σ in the transition region where the change of sign takes place and values between 1·5 and 2·5 are found.

§ 5. CONCLUSION

The general behaviour of the Hall coefficient and conductivity of lead sulphide, selenide and telluride can be accounted for by the conventional simple model of a semiconductor. A more detailed discussion of these results will be published subsequently, where it will be seen that the scope of the analysis is limited by the high concentration of extrinsic carriers found in most specimens. Further progress cannot be made until advances have been made in preparation and crystal growing techniques. It has been suggested recently (Goldberg and Mitchell 1954) that it may not be possible to grow stoichiometric crystals from the melt. It is therefore necessary to develop new techniques of crystal growth and treatment before crystals with smaller concentration of extrinsic carriers can be produced, enabling the measurement of the intrinsic properties to be made at lower temperatures than reported in this paper.

ACKNOWLEDGMENTS

The crystals used were supplied by W. D. Lawson and mounted in holders by W. H. Mitchell. Assistance with the measurements was given by J. B. Arthur and Miss M. Stallard.

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, Her Majesty's Stationery Office, for permission to publish this paper.

REFERENCES

BAUER, K., 1940, *Ann. Phys., Lpz.*, **38,** 84.

CHASMAR, R. P., and PUTLEY, E. H., 1951, *Semi-Conducting Materials*, ed. H. K. Henisch (London : Butterworth), p. 208.

DEVYATKOVA, E. D., MASLAKOVETS, Y. P., STILBANS, L. S., and STAVITSKAYA, T. S., 1952, *Dokl. Acad. Nauk S.S.S.R.*, **84,** 681.

EHRENBERG, W., and HIRSCH, J., 1951, *Proc. Phys. Soc.* B, **64,** 700.

EISENMANN, L., 1940, *Ann. Phys., Lpz.*, **38,** 121.

GOLDBERG, A. E., and MITCHELL, G. R., 1954, *J. Chem. Phys.*, **22,** 220.

HINTENBERGER, H., 1942, *Z. Phys.*, **119,** 1.

HOWARTH, D. J., and SONDHEIMER, E. H., 1953, *Proc. Roy. Soc.* A, **53,** 219.

LAWSON, W. D., 1951, *J. Appl. Phys.*, **22,** 1444; 1952, *Ibid.*, **23,** 495.

MITCHELL, W. H., 1954, *J. Sci. Instrum.*, **31,** 147.

PARKINSON, D. H., and QUARRINGTON, J. E., 1954, *Proc. Phys. Soc.* A, **67,** 569.

PETRITZ, R. L., SCANLON, W. W., and LUMMIS, F. L., 1954, *Bull. Amer. Phys. Soc.*, **29,** 32.

PUTLEY, E. H., 1952 a, *Proc. Phys. Soc.* B, **65,** 388; 1952 b, *Ibid.*, **65,** 991; 1952 c, *Ibid.*, **65,** 993; 1955, *Ibid.*, **68,** 35.

PUTLEY, E. H., and ARTHUR, J. B., 1951, *Proc. Phys. Soc.* B, **64,** 616.

SCANLON, W. W., 1953, *Phys. Rev.*, **92,** 1573.

Thermoelectric and Galvanomagnetic effects in Lead Selenide and Telluride

By E. H. PUTLEY
Radar Research Establishment, St. Andrew's Road, Malvern, Worcs.

MS. received 16*th August* 1954; *and in final form* 22*nd September* 1954

Abstract. Measurements of the thermoelectric power and the Peltier, Nernst, Ettingshausen and Righi–Leduc effects are described. It is shown that the behaviour of these effects is in accordance with the expectations of conventional semiconductor theory. The coefficients are also calculated from the known values of the Hall coefficient and conductivity of the specimen and good agreement is obtained between the measured and calculated galvanomagnetic coefficients. From observation of the Peltier effect an estimate of the thermal conductivity is obtained. Comparison of the measured and calculated values of the thermoelectric power enabled the effective mass of carriers in PbSe to be estimated.

---

§ 1. Introduction

An account has been given of measurements of the Hall coefficient and conductivity for single crystals of lead sulphide, selenide, and telluride (Putley 1955). In addition to these measurements, a number of other effects were measured on some specimens of lead selenide and lead telluride. The thermoelectric power and the Nernst effect were measured between 100 and 1000°K. Some measurements were made near room temperature of the Peltier, Ettingshausen and Righi–Leduc effects.

Since the Hall coefficient and conductivity of these specimens have been measured, it is possible to calculate the thermoelectric power and the other coefficients and to compare the calculated and measured values. It is found that the behaviour of the measured coefficients is in agreement with theory and the numerical values found for the galvanomagnetic coefficients are in reasonable agreement with the calculated values. By comparing the calculated and measured values of the thermoelectric power, an estimate for the effective mass of the current carriers can be obtained, and from observations of the Peltier effect the thermal conductivity was estimated.

§ 2. Methods of Measurement

The specimens used were prepared in the way described in the paper on Hall effect measurements (Putley 1955) but they were mounted in the crystal holder shown in figure 1. This holder is a modification of the one described for Hall effect measurements. A loop of 0·008 in. diameter platinum wire is welded at each end of the crystal and each end of the loop is taken to one of the four brass support wires. Thus, by suitable connections a current may be passed through the crystal between the loops or a current may be passed through either loop to heat one end of the crystal and thus set up a temperature difference between the ends of the crystal. By this means temperature differences up to 20–30° could be established. For measuring temperature four thermocouples were welded to points on the crystal. These thermocouples replaced the potential probes used in the simpler holder so that one pair of couples could be used to measure the potential difference between the ends of the crystal while the second pair could be used to measure transverse potential differences. With this arrangement both transverse and longitudinal differences of potential and of temperature could be measured simultaneously and thus the principal galvanomagnetic, thermomagnetic and thermoelectric effects could be studied.

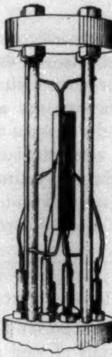


Figure 1. 12-pin crystal holder.

The couples attached to the crystals were made of 0·002 in. diameter wires of the NiCr–NiAl alloys T1 and T2 which were welded to the crystals using the technique described by Mitchell (1954). The holders could be used at all temperatures up to about 1000°K, but measurements of thermal effects were not made below about 100°K. The apparatus described in the preceding paper (Putley 1954) was used for varying the temperature of the crystals.

§ 3. Results

Figure 2 shows the variation with temperature of the thermoelectric power $d\theta/dT$ of three specimens of lead selenide. It is seen that at low temperatures two of the specimens were p-type and one n-type. As the temperature is increased,

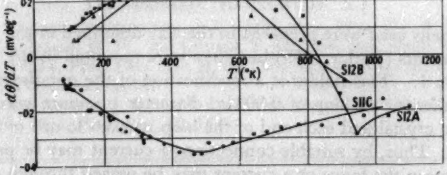


Figure 2. Thermoelectric power PbSe.

the magnitude of the thermoelectric power increases and passes through a maximum at about 500°K. This corresponds with the onset of mixed conduction in the specimen as will be seen by referring to Putley (1954) where the behaviour of the Hall coefficient R and conductivity σ of these specimens is described. Above 500°K the thermoelectric power decreases. In the two p-type specimens it changes sign near 800°K and at higher temperatures behaves in a similar way to the n-type specimen.

The Nernst coefficient Q was measured over a range of temperatures for PbSe specimen No. S12A. This effect was measured simultaneously with the thermoelectric power by measuring the transverse voltage across the crystal when a magnetic field was applied. The results obtained are shown in figure 3 together with the thermoelectric power, the Hall coefficient and the electrical conductivity for this specimen.

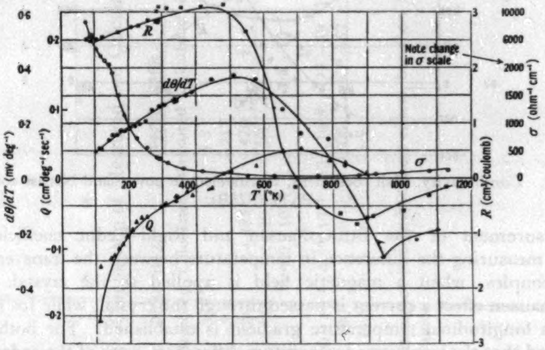


Figure 3. Conductivity, Hall coefficient, thermoelectric power and Nernst coefficient for PbSe (S12A).

Figure 4 shows the behaviour of the thermoelectric power, the Nernst and Hall coefficients and the electrical conductivity over a range of temperature for a specimen of PbTe. It is seen that the general behaviour of this specimen is similar to that found for the PbSe specimen. No measurements were made on PbS specimens, but the results obtained by Devyatkova *et al.* (1941) for PbS behave in a similar way to our measurements on PbSe and PbTe.

The Peltier effect was observed in a number of specimens in the following way. It was found that when a current of about 20 mA was passed through the crystal a difference in temperature of about 0·2° was observed between the pair of thermocouples mounted near each end of the crystal. When the current was reversed the temperature difference also reversed and therefore this temperature difference could be attributed to the Peltier heat liberated or absorbed at the current contacts. It is easy to show that the value of the temperature difference depends upon the ratio of the Peltier coefficient to the thermal conductivity. A cooling correction has to be applied to allow for heat loss in the current leads and by convection. This was found to be about 20%. Since the thermoelectric power could be measured, the Peltier coefficient could be determined using Thomson's relation $d\theta/dT = \pi/T$, and hence the thermal conductivity could be estimated. The values obtained for two specimens of PbSe and two of PbTe are given in table 1. In view of the uncertainty in estimating the cooling correction the values should be regarded only as approximate.

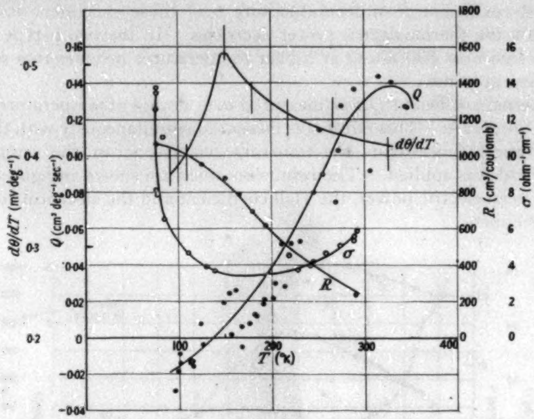


Figure 4. Conductivity, Hall coefficient, thermoelectric power and Nernst coefficient for PbTe (S28B).

Measurement of the Ettingshausen and Righi–Leduc coefficients both involve measuring the difference in temperature between the transverse pair of thermocouples, when a magnetic field is applied to the crystal. For the Ettingshausen effect a current is passed through the crystal, while for the Righi–Leduc a longitudinal temperature gradient is established. For both effects it was found that the transverse temperature differences were of the order of 0·01°. This was only just observable and was measured using a sensitive Kipp galvanometer which gave about 2 cm deflection per microvolt. The results of the measurements on two PbSe and two PbTe specimens are also given in table 1 which contains in addition the Nernst and Hall coefficients, the thermoelectric power and the electrical conductivity for these specimens, all measured at room temperature.

§ 4. THEORY OF THERMOELECTRIC AND GALVANOMAGNETIC EFFECTS IN SEMICONDUCTORS

Expressions for the various coefficients can be derived by applying the Lorentz–Sommerfeld theory as described by Seitz (1940). As the calculations are straightforward but tedious they will not be discussed. The expressions derived for the various effects are given in table 2 for both extrinsic and mixed semiconductors. In the discussion of our results the measured values will be compared with the values calculated from the formulae in table 2. Some of these formulae were first described by Bronstein (1932) and have been discussed recently (Putley 1952). The sign convention used is that given by Gerlach (1928). According to this the Hall effect, thermoelectric power and Righi–Leduc effect have a negative sign for n-type specimens and a positive sign for p-type.

Table 1. Thermo- and Galvanomagnetic Effects—Comparison of Measured and Calculated Values at Room Temperature

Specimen	PbSe S11C		PbSe S12A		PbTe S28B		PbTe S32B		
Effect	Measured	Calculated	Measured	Calculated	Measured	Calculated	Measured	Calculated	Units
Electrical conductivity σ	150		359		5·5		81·8		ohm^{-1} cm^{-1}
Hall coefficient R	−8		+2·9		−217		9·79		cm^{3} Coulomb^{-1}
Thermoelectric power 300°K	−290	−471	+260	+387	−420	†	+445	500	μv deg^{-1}
do. do. 100°K	−100	−308	+ 98	+230	−390				
Thermal conductivity		$4·2\times10^{-2}$		$4·0\times10^{-2}$		$3·8\times10^{-2}$		$6·3\times10^{-2}$	watt cm^{-1} deg^{-1}
(estimated from Peltier effect)						‡			
Nernst coefficient Q	$2·3\times10^{-2}$	$2·7\times10^{-2}$	$4·0\times10^{-2}$	$4·5\times10^{-2}$	13×10^{-2}	$QT=39$	$1·47\times10^{-2}$	$3·5\times10^{-2}$	cm^{2} sec^{-1} deg^{-1}
Ettingshausen coefficient	440	370	730	340	1500	$KP=57$	282	165	cm^{3} deg Joule^{-1}
Righi–Leduc coefficient	25	17	16·5	37	23·9		14	16	cm^{2} sec^{-1} volt^{-1}

† This is a mixed conductor. Have not attempted to calculate these effects, but show Bridgman relation obeyed.
‡ Bridgman relation $QT=KP$ numerical values.

Table 2. Formulae derived from the Lorentz–Sommerfeld Theory

List of symbols

n	concentration of conduction electrons
μ_1	mobility of conduction electrons
l_1	mean free path of electrons
m_1	effective mass of electrons
K_L	contribution of crystal lattice to the thermal conductivity
E	energy gap between the bands
ϵ	chemical potential (i.e. for classical semiconductors $\epsilon = \boldsymbol{k}T \ln\left(\frac{nh^3}{2(2\pi m_1 \boldsymbol{k}T)^{3/2}}\right)$
ξ	chemical potential for positive holes (i.e. $\xi = -E - \epsilon$) $= \boldsymbol{k}T \ln\left(\frac{ph^3}{2(2\pi m_2 \boldsymbol{k}T)^{3/2}}\right)$
T	absolute temperature
$E/\boldsymbol{k}T$	is written E'
p	concentration of holes in full band
μ_2	mobility of holes
l_2	mean free path of holes
m_2	effective mass of positive holes
σ	electrical conductivity
K	thermal conductivity
R	isothermal Hall effect
Q	isothermal Nernst effect
P	Ettingshausen effect
S	Righi–Leduc effect
$\frac{d\theta}{dT}$	thermoelectric power

	Extrinsic Conductors		Mixed conductors
Effect	n-Type	p-Type	
σ	$\frac{4}{3}\frac{n\boldsymbol{e}^2 l_1}{(2\pi m_1 \boldsymbol{k}T)^{1/2}} = n\boldsymbol{e}\mu_1$	$\frac{4}{3}\frac{p\boldsymbol{e}^2 l_2}{(2\pi m_2 \boldsymbol{k}T)^{1/2}} = p\boldsymbol{e}\mu_2$	$n\boldsymbol{e}\mu_1 + p\boldsymbol{e}\mu_2$
K	$K_L + 2\frac{\boldsymbol{k}^2T}{\boldsymbol{e}} n\mu_1$	$K_L + 2\frac{\boldsymbol{k}^2T}{\boldsymbol{e}} p\mu_2$	$K_L + \frac{\boldsymbol{k}^2T}{\boldsymbol{e}} \frac{2(n\mu_1+p\mu_2)^2 + np\mu_1\mu_2(4+E')^2}{n\mu_1+p\mu_2}$
R	$-\frac{3\pi}{8}\frac{1}{n\boldsymbol{e}}$	$+\frac{3\pi}{8}\frac{1}{p\boldsymbol{e}}$	$-\frac{3\pi}{8\boldsymbol{e}}\frac{n\mu_1^2 - p\mu_2^2}{(n\mu_1+p\mu_2)^2}$
Q	$-\frac{3\pi}{16}\frac{\boldsymbol{k}}{\boldsymbol{e}}\mu_1$	$-\frac{3\pi}{16}\frac{\boldsymbol{k}}{\boldsymbol{e}}\mu_2$	$-\frac{3\pi}{16}\frac{\boldsymbol{k}}{\boldsymbol{e}}\frac{(n^2\mu_1^3+p^2\mu_2^3) - np\mu_1\mu_2(\mu_1+\mu_2)(7+2E')}{(n\mu_1+p\mu_2)^2}$
P	$-\frac{3\pi}{16}\frac{\boldsymbol{k}T}{\boldsymbol{e}}\frac{\mu_1}{K}$	$-\frac{3\pi}{16}\frac{\boldsymbol{k}T}{\boldsymbol{e}}\frac{\mu_2}{K}$	$-\frac{3\pi}{16}\frac{\boldsymbol{k}T}{eK}\frac{(n^2\mu_1^3+p^2\mu_2^3) - np\mu_1\mu_2(\mu_1+\mu_2)(7+2E')}{(n\mu_1+p\mu_2)^2}$
S	$-\frac{21\pi}{32}\frac{\boldsymbol{k}^2T}{\boldsymbol{e}K} n\mu_1^2$	$+\frac{21\pi}{32}\frac{\boldsymbol{k}^2T}{\boldsymbol{e}K} p\mu_2^2$	$+\frac{3\pi}{32}\frac{\boldsymbol{k}^2T}{\boldsymbol{e}K}\frac{7(p\mu_2^2 - n\mu_1^2)(n\mu_1+p\mu_2)^2 + 4np\mu_1\mu_2\{(\mu_1-\mu_2)(n\mu_1+p\mu_2)(4+E') + \mu_1\mu_2(n-p)(4+E')^2\}}{(n\mu_1+p\mu_2)^2}$
$\frac{d\theta}{dT}$	$-\frac{\boldsymbol{k}}{\boldsymbol{e}}\left(2 - \frac{\epsilon}{\boldsymbol{k}T}\right)$	$+\frac{\boldsymbol{k}}{\boldsymbol{e}}\left(2 - \frac{\xi}{\boldsymbol{k}T}\right)$	$-\frac{\boldsymbol{k}}{\boldsymbol{e}}\frac{n\mu_1(2-\epsilon/\boldsymbol{k}T) - p\mu_2(2-\xi/\boldsymbol{k}T)}{n\mu_1+p\mu_2}$

As is well known, in a p-type specimen these effects can become negative at high temperatures when intrinsic conduction becomes dominant. The Nernst and Ettingshausen effects have a negative sign for both n- and p-type extrinsic conductors, but they may change sign if mixed conduction is present. Thus referring to the expressions in table 2 it is seen that these effects will change sign if

$$\frac{n^2\mu_1{}^3+p^2\mu_2{}^3}{np\mu_1\mu_2(\mu_1+\mu_2)}<7+\frac{2E}{kT},$$

and if $n=p$ this condition will be satisfied provided $\mu_1<8\mu_2$, a condition which applies to all known intrinsic semiconductors except indium antimonide. Thus in a p-type semiconductor the effects should change sign as the temperature is raised to the intrinsic region in the following order: first Ettingshausen and Nernst effects, next the Hall effect then the thermoelectric power and the Righi–Leduc effect last. In an n-type specimen the Nernst and Ettingshausen effects will change sign when intrinsic range is approached, but the others will not.

§ 5. DISCUSSION

If the Hall coefficient and conductivity are known, it is possible to calculate the other coefficients and compare the calculated and measured values. It is found that there is broad agreement between experiment and calculation based on classical semiconductor theory.

According to theory, the Nernst effect should have a negative sign in an extrinsic conductor, but it should change sign when mixed conduction sets in. Referring to figures 2 and 4 it is seen that exactly this took place. The PbSe specimen was p-type and the PbTe was n-type, but they both had a negative Nernst coefficient at lower temperatures which became positive as the intrinsic range was approached. Thus, by observing the sign of the Nernst effect, it can be seen whether mixed conduction is present or not.

In a p-type semiconductor both the Hall coefficient and thermoelectric power will change sign when mixed conduction sets in. For the PbSe specimen shown in figure 3 as the temperature increased the Nernst coefficient changed sign first, the Hall coefficient next and finally the thermoelectric power. The temperature at which the thermoelectric power is zero is approximately equal to that at which the Hall coefficient passes through its minimum value. This behaviour is in agreement with the theory.

The Nernst, Ettingshausen and Righi–Leduc coefficients for an extrinsic conductor may be calculated from the Hall coefficient and conductivity using the formulae given in table 2. The results of these calculations are given in table 1 and it is seen that there is reasonable agreement between the measured and calculated values. The thermoelectric power may also be calculated from the Hall coefficient using the formulae given in table 2. Values calculated from this expression are given in table 1. In this calculation m was taken equal to the free electron mass and it is seen that there is a discrepancy between the calculated and measured values which is too large to be accounted for by experimental errors. By comparing the calculated and measured values the effective mass of the current carriers can be estimated. Table 3 gives the values found for the effective mass for three specimens of PbSe. It is seen that values in the region of $0{\cdot}3m$ are obtained.

Having determined the Ettingshausen coefficient and the thermoelectric power, it is possible to estimate the error which may occur in d.c. Hall effect measurements by the presence of a Seebeck e.m.f. produced by the Ettingshausen temperature difference. If the temperature difference is about 0·01°K and the thermoelectric power about 0·4 mv $\deg^{-1}$ this e.m.f. will be about 4 μv. Since the Hall voltages are usually several hundred microvolts it is seen that this error will normally be less than about 1%.

Table 3. Effective Mass estimated from Thermoelectric Power for PbSe

Specimen	T (°K)	Thermoelectric power (μv $\deg^{-1}$) Measured	Thermoelectric power (μv $\deg^{-1}$) Calculated	Effective mass / Free mass
S11C	300	−290	−471	0·25
	100	−100	−308	0·20
S12A	300	+260	+387	0·36
	100	+ 98	+230	0·36
S12B	300	+210	+396	0·24
	100	+ 60	+245	0·24

Bronstein's formula shows that the Nernst coefficient is proportional to the product $R\sigma$ and therefore it should vary with temperature in the same way as the mobility. If the results shown in figure 3 are re-plotted on a log–log scale it is found that the Nernst coefficient varies as $T^{-5/2}$, which is in agreement with the behaviour found for the $R\sigma$ product (Putley 1954).

§ 6. CONCLUSION

The general behaviour of the effects studied can be accounted for by conventional semiconductor theory. The effective mass of electrons in PbSe can be estimated from measurement of the thermoelectric power and is found to be about 0·3m. The thermal conductivity can be estimated from the Peltier temperature difference set up when a current is passed through the crystal. The values obtained for the Ettingshausen coefficient shows that errors in d.c. Hall effect measurements due to this effect should not be greater than 1% which is not large compared with other errors that are likely to be present.

ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, Her Majesty's Stationery Office, for permission to publish this paper.

REFERENCES

BRONSTEIN, M., 1932, *Phys. Z. Sowjet.*, **2**, 28.

DEVYATKOVA, E. D., MASLAKOVETS, U. P., and SOMINSKII, M. S., 1941, *Izv. Akad. Nauk., S.S.S.R.*, **5**, 409.

GERLACH, W., 1928, *Handbuch der Physik*, **13** (Berlin : Julius Springer), p. 228.

MITCHELL, W. H., 1954, *J. Sci. Instrum.*, **31**, 147.

PUTLEY, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 991; 1955, *Ibid.*, **68**, 22.

SEITZ, F., 1940, *The Modern Theory of Solids* (New York : McGraw-Hill).

Carrier Extraction in Germanium

By J. B. ARTHUR, W. BARDSLEY, M. A. C. S. BROWN AND A. F. GIBSON
Radar Research Establishment, Great Malvern, Worcs.

MS. received 25*th May* 1954 *and in final form* 27*th October* 1954

Abstract. It is well known that the density of current carriers (electrons and holes) in a germanium crystal can be increased by injection from a rectifying contact. The purpose of this paper is to show that large changes in carrier concentration can be obtained in near-intrinsic germanium by the reverse of injection, namely extraction. This technique allows some new fundamental experiments to be made and, by way of example, an experiment on the drift mobility of carriers in intrinsic germanium is described.

§ 1. INTRODUCTION

It is well known that carriers may be injected into germanium from a rectifying contact. This phenomenon forms the basis of transistor action. It is not, however, so widely appreciated that the reverse of injection, namely extraction, must also exist. Banbury (1953) noticed a small drop in the conductance of collector contacts on PbS filaments when a sweeping field was applied and ascribed this effect to extraction. Some discrepancies between theory and experiment noticed by Dacey and Ross (1953) in unipolar transistors have also been ascribed to extraction effects. It is our purpose to show that quite large changes in carrier concentration can be obtained by extraction methods in suitable germanium crystals. Proportionately much larger changes can be obtained by extraction than by injection, thus allowing some new and potentially valuable devices to be made.

§ 2. THE METHOD OF EXTRACTION

Consider a rod of n-type germanium containing n_e electrons per cm^3 due to impurities (assumed to be all ionized). The total density of electrons will then be (n_e+p_0) if the density of holes is p_0. If an electric field is now applied to the rod the holes will be swept along the rod and out through the negative connection. From space charge considerations an equal number of electrons must also be removed. If, therefore, the rate of generation of holes within the material plus the rate of injection of holes from the positive connection are together less than the rate of removal of holes by the field the density of holes will be reduced and extraction may be said to have occurred.

To put the above argument on a semi-quantitative basis suppose that the rate of hole injection is small compared with the rate of generation. The techniques that may be used to achieve this condition are described in a later section. Let p be the density of holes when some extraction has taken place, where p is less than p_0, the equilibrium density. Then the rate of generation of holes is given by p_0/τ, the rate of recombination of holes by p/τ and the rate of extraction of holes by p/t where τ is the minority carrier (hole) lifetime and t is the transit time of holes along the crystal.

In equilibrium the rates of recombination and extraction must together equal the rate of generation, so that

$$\frac{p_0}{\tau}=p\left(\frac{1}{\tau}+\frac{1}{t}\right) \text{ or } p=p_0\left(\frac{t}{\tau+t}\right). \qquad \ldots\ldots(1)$$

Thus if sufficiently large voltages are applied, such that $t \ll \tau$, a noticeable amount of extraction should occur. The above equation (1) should, however, be considered as indicative rather than accurate. It is over-simplified in the following respects:

(i) The generation and recombination rates used are valid only when $p \ll (p+n_e)$. In near-intrinsic germanium this condition is not obtained except when an appreciable amount of extraction has occurred ($t \ll \tau$).

(ii) The hole density is assumed to be uniform throughout the specimen. This is valid only for $t=\infty$. Otherwise p is of the form $p=p_0\,\{1-\exp(t'/\tau)\}$ where t' is the transit time of a hole from the positive end of the crystal to any point in it. For $t \ll \tau$ this reduces to $p=p_0\,t'/\tau$ and p in equation (1) may be interpreted as the maximum hole density, which occurs at the negative extraction electrode ($t=t'$).

(iii) The lifetime is assumed to be invariant with hole density. The theory of Shockley and Read (1952) indicates that the hole lifetime in relatively pure germanium is given by

$$\tau=\tau_0\left(\frac{n_e+2p_0}{n_e+p+p_0}\right).$$

An increase in lifetime under extraction conditions can therefore be expected. This, of course, makes the condition $t<\tau$ easier to achieve, but, at most, the above effect only changes τ by a factor of 2.

Finally it is perhaps important at this juncture to point out the distinction between extraction as discussed in this paper and the depletion of minority carriers that occurs around a reverse biased rectifying contact. The latter is essentially a local effect, limited to a region within a diffusion length of the contact. The extraction effect discussed here extends throughout the specimen. This distinction between local depletion and extraction can probably be illustrated by the following example. In a depletion layer the density of minority carriers (holes) is at a minimum at the extracting (negatively biased) electrode and increases to the limiting undisturbed value at large distances from the extracting electrode. In an extracted rod, on the other hand the minority carrier density is a maximum at the negatively biased extraction electrode (see note (ii) above) and a minimum at the other (positively biased) electrode. In an extracted rod the minority carrier density is everywhere less than the original undisturbed density.

§ 3. THE CONDUCTIVITY OF EXTRACTED GERMANIUM

At zero or very low voltages such that $t>\tau$, the conductivity of the sample is given by

$$\sigma=p_0 q\mu_p+(p_0+n_e)q\mu_n$$

and under extraction conditions ($t<\tau$) the conductivity is given by

$$\sigma_e=pq\mu_p+(p+n_e)q\mu_n$$

where q is the electronic charge and μ_p and μ_n the hole and electron mobilities respectively. It will be seen that the conductivity will be reduced on extraction to reach a limiting value of $\sigma_e \simeq n_e q \mu_n$ for small values of p. The actual value of p is then no longer important. In practice, therefore, the resistance of the sample will reach some limiting value determined by the density of impurities. In physical terms the impurity electrons cannot be extracted for, if they were, the crystal would carry a net positive charge due to the ionized impurities remaining in the crystal.

§ 4. PRACTICAL CONSIDERATIONS

Equation (1) above assumes that the only source of holes is thermal generation in the specimen. In practice, however, an appreciable number of minority carriers may be injected into the specimen from the positive end connection. If the rate of injection is comparable with the rate of thermal generation of carriers equation (1) no longer applies and extraction is not likely to be observed. The importance of injection from the end connections can best be illustrated by means of the following example.

If the current through the specimen is i amps, the injection ratio (Shockley, Pearson and Haynes 1949) of the end connection γ, and V the volume of the crystal we require the condition

$$i\gamma/q \leqslant pV/10\tau \qquad \ldots\ldots(2)$$

where the factor 10 has been introduced simply to ensure that the rate of carrier injection is at least an order of magnitude less than the generation rate. For a typical specimen having dimensions of 2 mm × 0·5 mm × 0·5 mm, a lifetime of 100 μsec and requiring a current of 1 mA, equation (2) requires $\gamma \leqslant 5 \times 10^{-4}$. The injection ratio of simple mechanical contacts on germanium is known to be about unity and for soldered (apparently ohmic) connections γ lies between 10^{-2} and 10^{-1} (Many 1954). It can be shown (unpublished calculations) that 'dumb-bell' ends (Bond 1950) are of no value in reducing injection in near-intrinsic material. Thus some consideration must be given to the end connections if equation (2) is to be satisfied.

Consider, then, a sample of near-intrinsic but slightly n-type germanium. The sample being n-type we need only consider the injection of holes from the positively biased end. If this end is now made strongly n-type by heavy doping (see figure 1) three results are obtained, namely: (i) The soldered end connection can be placed sufficiently far from the near-intrinsic section so that all the injected holes have recombined before reaching the relevant section. (ii) The density of holes in the heavily doped section can be made so small that most of the current from the heavily doped to the near-intrinsic section is carried by electrons. (iii) The lifetime in the heavily doped section can be made sufficiently long to ensure a small rate of generation of holes in this section, as these holes enter the

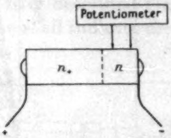


Figure 1.

near-intrinsic section by diffusion. It follows, therefore, that the problem of making a suitable non-injecting end connection reduces to a crystal growing problem. There are two ways in which a heavily doped end may be made: (*a*) Direct growth from the melt by well-known crystal pulling methods. The crystals used for the experiments described in this paper were prepared in this way. A number of crystals of various resistivities have been grown, the doping agent in the heavily doped section being antimony in all cases. The characteristics of a typical specimen were for near intrinsic section, resistivity 40 ohm cm, hole lifetime (in filament) 100 μsec, and for heavily doped section, resistivity 0·002 ohm cm, hole lifetime 40 μsec. (*b*) The alloying method: while this procedure is not amenable to quantitative control it has been found that alloying with a Pb–Sb alloy, following N–P–N junction techniques, produces suitable end connections.

§ 5. EXTRACTION BY D.C. FIELDS

To determine the degree of carrier extraction occurring in a germanium crystal of the above specification the resistivity of the high resistivity section is determined potentiometrically, as indicated in figure 1. This procedure eliminates the resistance, if any, of the N–N junction, end connections, etc. and was adopted for this reason but it has been found in practice that the resistance of the whole assembly is almost entirely determined by the high resistivity section.

The current–voltage characteristics of a typical specimen at various temperatures are shown in figure 2. These results were obtained in the extraction direction, namely with the heavily doped end made positive. When the polarity is reversed the resistance of the crystal usually decreases (due to injection) but sometimes increases if the soldered connection is sufficiently good to allow extraction to occur. Some units have been made with heavily doped sections at both ends, when symmetrical curves are obtained.

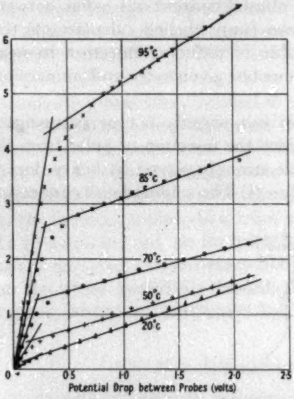


Figure 2.

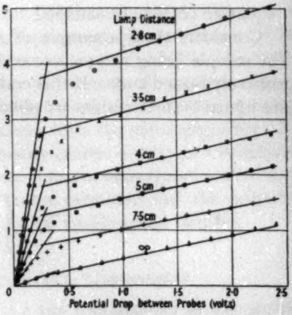


Figure 3.

Inspection of figure 2 shows that the expected results are obtained, at least qualitatively. The initial slope of each curve corresponds to the unextracted conductivity and this increases rapidly with temperature, as would be expected for a near-intrinsic specimen. The final slope, corresponding to the conductivity due to the original impurities, should be substantially invariant with temperature as the mobility only varies slowly. This result is also obtained.

In figure 3 are shown a similar set of curves for the same specimen obtained at room temperature under various levels of illumination. The light source was a 36-watt tungsten filament lamp, the distances between lamp and sample being indicated on the diagram. As might be expected, the effects of light and heat are essentially similar in producing minority carriers and figures 2 and 3 are likewise similar.

It will be seen from figures 2 and 3 that the voltage required to produce a significant amount of extraction (referred to as the extraction voltage) can be defined fairly accurately. If the extraction voltage is determined by extrapolation of the unextracted and extracted portions it is found to be about 0·3 volt for the sample quoted and largely invariant with light intensity (i.e. hole density). Figure 2 shows, however, that the extraction voltage increases slowly with increasing temperature.

For the reasons indicated in § 2 it is not possible to calculate the extraction voltage exactly, particularly in almost intrinsic material. For the purpose of designing devices based on extraction, however, it has been found sufficient to assume the validity of equation (1) for all values of t and to suppose that a significant amount of extraction will occur when $t=\tau$. Using this condition and writing the familiar relationships $t=l^2/\mu_p V_E$ and $L_p^2=\mu_p \boldsymbol{k} T\tau/q$ we obtain

$$V_E=\frac{\boldsymbol{k}T}{q}\left(\frac{l}{L_p}\right)^2 \qquad \ldots\ldots(3)$$

where $\boldsymbol{k}$ is Boltzmann's constant, T the absolute temperature, l the length of the specimen and V_E the extraction voltage. Equation (3) has been found to give a very good indication of the value of V_E for values of V_E between 0·1 and 10·0 volts for samples ranging in resistivity from 8 ohm cm to 50 ohm cm at room temperature.

§ 6. EXTRACTION BY PULSED FIELDS

In the previous section we have described the static characteristics of extraction systems; in this section we shall indicate their dynamic characteristics.

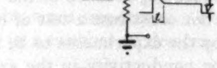


Figure 4.

If a voltage pulse of suitable magnitude is applied to the crystal extraction will proceed during the period of the pulse until the limiting resistance is reached. The output waveform across a suitable load resistance will, therefore, be of the form shown in figure 4. That this type of waveform is, in fact, observed is in itself a qualitative indication that extraction is taking place. The time t taken for extraction is determined by the mobility of the minority carriers (holes) in the near-intrinsic section of the crystal. This time can be measured with considerable accuracy as a very sharp 'knee' occurs at this time. The sharp knee is a consequence of the shock wave that builds up on the trailing edge of the carrier departure wave. This is a well-known effect in drift mobility experiments and is due to the large change in carrier concentration at the edge of the carrier departure wave (see Blakemore *et al.* 1953 for references).

In carrying out this experiment two points must be noted: (i) The extraction field in the filament must be constant with time, at least within the accuracy of field measurement, about 5%. A linear change in output voltage with time during extraction (figure 4) implies a uniform velocity and has been used as a criterion of constant field conditions.

(ii) The velocity of the shock wave measured is not exactly equal to the velocity of the disturbance. The magnitude of this effect has been calculated by J. B. Gunn (private communication). The two velocities are, of course, the same when $\sigma/\sigma_e = 1$ and differ by 2% for $\sigma/\sigma_e = 5$ and by 4% for $\sigma/\sigma_e = 10$. This effect has, therefore, been ignored.

If the extraction time t is measured the carrier velocity and drift mobility may be determined. This experiment, moreover, allows the drift velocity of carriers to be determined well into the intrinsic region of conductivity, i.e. when the numbers of holes and electrons are comparable. For a variety of reasons the drift mobility of injected carriers cannot be studied experimentally in this region. It is expected theoretically that the group velocity of carriers will fall to zero when the semiconductor becomes very nearly intrinsic (Shockley 1950) the appropriate relation being quoted by Prince (1953 a) as

$$\mu_g = \mu \left[1 + \left(1 + \frac{1}{b} \right) \frac{p}{n_e} \right]^{-1} \qquad \ldots\ldots(4)$$

for n-type material.

Here μ_g is the effective group mobility of holes, μ the 'normal' undisturbed mobility of holes and b the ratio of electron to hole mobility. It will be noticed that, for normally doped samples ($n_e \gg p$), $\mu_g = \mu$ and for intrinsic samples ($n_e = 0$), $\mu_g = 0$.

As far as we are aware, the validity of equation (4) has never been checked directly, due presumably to the difficulty of making drift measurements in this region. Equation (4) has been applied by Prince (1953 a) to explain the deviation at high temperatures, observed by himself and by Lawrance (1953), from the simple power relationship between hole mobility and temperature. To obtain the appropriate correction factor it is necessary to assume the value of b and the absolute value and temperature dependence of the density of intrinsic carriers. Hence this treatment does not constitute a test of equation (4). The equation can, however, be checked by the experiments to be described.

Referring to figure 3 the conductivity in the extracted condition σ_e can be written as

$$\sigma_e = n_e q \mu b. \qquad \ldots\ldots(5)$$

The conductivity in the unextracted condition σ can be written as

$$\sigma = \sigma_e + p q \mu (1 + b) \qquad \ldots\ldots(6)$$

where μ is the normal mobility of holes. Combining equations (4), (5) and (6) we find

$$\mu_g/\mu = \sigma_e/\sigma \qquad \ldots\ldots(7)$$

a relation which can be checked directly by experiment. The carrier group mobility μ_g, together with σ_e and σ, is determined at room temperature under various levels of illumination and a value of μ obtained. This may then be compared with the known value of $1850\ \mathrm{cm^2\,v^{-1}\,sec^{-1}}$ appropriate to these conditions (Prince 1953 b).

In figure 5 we show the carrier group velocity as a function of sweeping field under various levels of illumination. The fall in group velocity with increasing light intensity is evident, the slope of each line representing the mobility. The values of σ_e and σ can then be determined for each light intensity and the ratio σ_e/σ plotted against the effective group mobility μ_g. This is done in figure 6. The resulting straight line must pass through the origin and the best line indicates a value of μ of $1840\ \mathrm{cm^2\,v^{-1}\,sec^{-1}}$ in excellent agreement with expectations. The experiment therefore confirms the validity of equation (4).

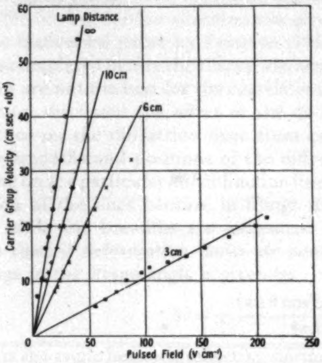


Figure 5.

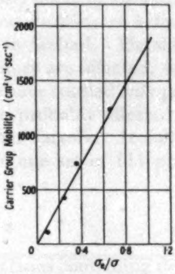


Figure 6.

§ 7. CONCLUSIONS

It is believed that the existence of extraction has been adequately demonstrated. Some indication of the new experiments that may be done using extraction has been given. In addition, extraction techniques can increase the temperature at which intrinsic conduction becomes important. This feature makes the effect have important applications in a number of devices.

ACKNOWLEDGMENTS

We are indebted to our colleagues in the transistor group of this Establishment for their advice and help. The paper is published by permission of the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office.

REFERENCES

BANBURY, P. C., 1953, *Proc. Phys. Soc.* B, **66,** 50.

BLAKEMORE, J. S., DE BARR, A. E., and GUNN, J. B., 1953, *Rep. Progr. Phys.* (London : Physical Society), **16,** 160.

BOND, W. L., 1950, *Phys. Rev.*, **78,** 646.

DACEY, G., and ROSS, I. M., 1953, *Proc. Inst. Radio Engrs, N.Y.*, **41,** 970.

LAWRANCE, R., 1953, *Phys. Rev.*, **89,** 1295.

MANY, A., 1954, *Proc. Phys. Soc.* B, **67,** 9.

PRINCE, M. B., 1953 a, *Phys. Rev.*, **91,** 271 ; 1953 b, *Ibid.*, **92,** 681.

SHOCKLEY, W., 1950, *Electrons and Holes in Semi-conductors* (New York : Van Nostrand), p. 329 *et seq.*

SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R., 1949, *Bell. Syst. Tech. J.*, **28,** 344.

SHOCKLEY, W., and READ, W., 1952, *Phys. Rev.*, **87,** 835.

RESEARCH NOTES

Deformation Faults in Cold-Worked Metals

By G. B. GREENOUGH† and E. M. SMITH‡
Royal Aircraft Establishment, Farnborough, Hants.

MS. received 4*th October* 1954

THE experimental work on which this note is based was carried out during 1950–51, but further attention was given to the results after the appearance of a theoretical paper by Paterson (1952). The fact that the experimental work was completed before the theory was available explains why the experimental conditions are not the best for the correlation of theory and experiment.

Paterson discussed the effect of the deformation faulting of a face-centred cubic lattice on the diffraction lines from crystals so faulted. He showed that both the breadths and positions of the diffraction lines are affected, to a degree depending on the particular *hkl* diffraction line. We have studied only the changes in position of the lines because in filings these are probably affected solely by faulting while the breadths are influenced by other factors. It follows from Paterson that, if deformation faults are confined to one set of 111 planes only, the change in the Bragg angle is given by

$$\delta\theta = \pm \frac{\tan\theta \cos^2\phi \, 3\sqrt{3}}{4\pi h_3}\alpha \qquad \ldots\ldots(1)$$

where ϕ is the angle between reflecting normal and planes containing deformation faults, α is the probability of a stacking fault and $h_3 = |h+k+l|$. This formula applies only to x-ray reflections for which $h_3 = 3n \pm 1$ (n an integer), being positive or negative according as h_3 is $3n+1$ or $3n-1$. When $h_3 = 3n$, $\delta\theta = 0$.

Two complicating features arise when powders are used in the investigation. The first, that the diffraction lines are the combined reflections from many planes (hkl, $hk\bar{l}$ etc.), is easily accommodated by taking the appropriate mean value for $\cos^2\phi$. The second is that stacking faults probably exist in the cold-worked metal parallel to all planes of the form {111}. Diffraction from such a lattice has not been examined theoretically and we have assumed that the effects of faults parallel to all planes {111} on the diffraction line positions are linearly additive.

The first material examined was O.F.H.C. copper. Filings were made from annealed rod and diffraction photographs taken of those of sizes between 250 and 200 mesh using a Bradley–Jay 19 cm camera and CuKα radiation. The films were then photometered and the position of the Kα_1 peak determined using the method described by Rachinger (1948). A similar procedure was employed to determine the positions of the diffraction line peaks for annealed copper filings. Six different specimens were examined in each case. The experimental scatter of the results was such that no significant difference could be detected between the individual θ values from the annealed and the cold-worked copper. However, since many camera errors are much reduced by considering the difference of θ

---

† Now at Department of Metallurgy, University of Sheffield.

‡ Address now ‘ Windrush ’, Ashley Road, Farnborough.

between neighbouring diffraction lines, the results for lines 19 and 20 were expressed:

$$\begin{array}{ll} \text{Annealed copper,} & \theta_{20}-\theta_{19}=4{\cdot}094^\circ \pm 0{\cdot}001 \\ \text{Cold-worked copper,} & \theta_{20}-\theta_{19}=4{\cdot}079^\circ \pm 0{\cdot}005 \end{array} \Bigg\} \text{(50\% confidence limits)}$$

Student's '*t*' test indicates that this difference is just significant (chance probability 0·03). It should also be noted that the lines have approached each other, agreeing with the theory of deformation faults. This difference could be caused by an increase in lattice parameter which, as well as producing a decrease in θ for all diffraction lines, would bring lines 19 and 20 closer together. A fractional increase in parameter of 4×10^{-4} would be required to explain the results quoted, but a further study of all the observations showed that any change in parameter was less than 5×10^{-5}. It is considered that the results demonstrate that stacking faults are present in cold-worked copper.

By substituting the appropriate results in equation (1), a value of $\alpha = 7 \times 10^{-3}$ is obtained. Although the limits of error for this value are very wide, it is significantly less than the figure $2{\cdot}5 \times 10^{-2}$ found by Warren and Warekois (1953) for cold-worked α-brass employing a similar technique. If it is assumed that the faults are equally distributed among the four possible {111}, then in a particular [111] in the copper filings there is one fault every 600 atom planes.

Similar work was also carried out on an aluminium alloy containing 8% silver after a heat-treatment of 24 hours at 300°c to precipitate some of the silver as $AlAg_2$. An alloy of aluminium was used since it recovers more slowly from the effects of cold-work than does pure aluminium at room temperature, and a silver alloy chosen in particular because silver in solution in aluminium has no effect on the lattice parameter of the latter (Axon and Hume-Rothery 1948). The results for this alloy were more accurate than for the copper since the line-broadening was less. Thus, instead of using the difference of the values for neighbouring lines, it is possible to examine the behaviour of each line separately. Lines 16, 19 and 20 all showed significant changes, viz.

Line	16	19	20
$\theta_{\text{ann.}} - \theta_{\text{c.w.}}$ (deg)	$-0{\cdot}042 \pm 0{\cdot}005$	$-0{\cdot}046 \pm 0{\cdot}003$	$-0{\cdot}067 \pm 0{\cdot}004$

These results cannot be explained entirely in terms of deformation faults in the lattice. However, they fit with surprising accuracy the postulate that, in addition to producing deformation faults, cold-working the alloy causes a change of the true lattice parameter. The numerical values are that the cold-working causes a decrease in lattice parameter $\delta a/a = 2 \times 10^{-4}$ and a density of stacking faults $\alpha = 4 \times 10^{-3}$. This last is the same, within the limits of error, as the value found for copper, and was expected. No satisfactory explanation can be offered for the decrease in parameter, although it may be associated with the removal by cold-working of the coherency strain between the lattices of the aluminium and $AlAg_2$ precipitate.

This note is published with the permission of the Chief Scientist, Ministry of Supply.

REFERENCES

AXON, H. J., and HUME-ROTHERY, W., 1948, *Proc. Roy. Soc.* A, **193,** 1.
PATERSON, M., 1952, *J. Appl. Phys.*, **23,** 805.
RACHINGER, W. A., 1948, *J. Sci. Instrum.*, **25,** 254.
WARREN, B. E., and WAREKOIS, E. P., 1953, *J. Appl. Phys.*, **24,** 951.

LETTERS TO THE EDITOR

The Infra-Red Luminescence of Solid Halogens

Investigation of luminescence in monatomic solids has revealed that the halogens chlorine, bromine and iodine and their interhalogen compounds show luminescence in the near infra-red when suitably excited in the solid state. The solidified specimens were excited by light from a tungsten lamp and their excitation and emission spectra analysed with a double prism monochromator (LiF prism), lead sulphide detector cell and tuned amplifier (800 c/s), the radiation being suitably chopped. For measurement of excitation spectra the monochromator is used with light source to provide excitation of selected wavelength. The total emission of the halogen is then detected by using suitable optical filters in front of the lead sulphide cell and radiation chopper which exclude the exciting radiation but pass the emission band. Excitation spectra are corrected to an equal energy basis for the excitation but are uncorrected for absorption change with wavelength in the halogen crystals.

The emission spectra of chlorine, bromine and iodine at 90°K are given in figure 1 which includes the excitation spectra for bromine and iodine. They consist of broad bands with no sub-structure evident. Experimental difficulties made it impossible to obtain more than a qualitative idea of the excitation region for chlorine. When plotted on an energy or wave number abscissa the emission curves assume a symmetrical form. The emission and excitation maxima move to longer wavelengths with increasing atomic number as shown. The emission

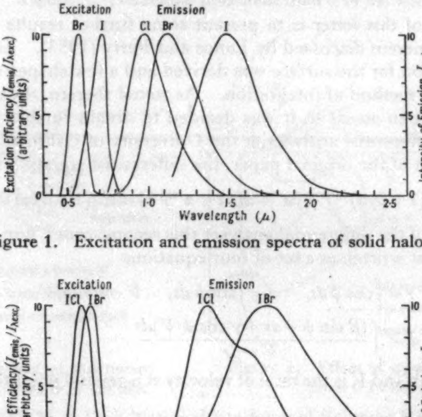


Figure 1. Excitation and emission spectra of solid halogens.

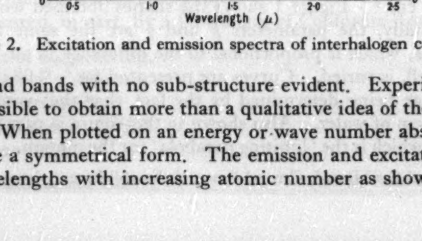


Figure 2. Excitation and emission spectra of interhalogen compounds.

in each case increases with decreasing temperature but quantitative measurements are still in progress. In the case of the emission of the interhalogen compounds at 90°K shown in figure 2 the maximum also moves to longer wavelengths with increasing atomic number, but comparison with figure 1 shows that for iodine bromide the peak emission is at a longer wavelength than that for iodine. The secondary peak in the emission band of iodine chloride is most probably due to a separate phase of iodine in the polycrystalline mass.

With further experimental information on such characteristics as polarization of luminescence, lifetimes of excited states and temperature dependence it should be possible to construct a model for the emission centres. These luminescence investigations may thus be valuable in helping towards an understanding of electronic levels in molecular crystals.

Luminescence Laboratory,
Department of Physics,
University of Birmingham.
19th October 1954.

M. J. Dumbleton†.

On the Differential Analyser Solution of the Water Bells Problem

The purpose of this letter is to present some further results concerning the water bell phenomenon discussed by Lance and Perry (1953). In this paper the differential equation for the surface was derived and a few shapes were presented, using a numerical method of integration. As stated therein, the accuracy of this method is difficult to assess so it was decided to obtain further solutions using the mechanical differential analyser at the University of California, Los Angeles.

In the notation of the original paper, the differential equation to be solved is

$$\{1+z'^2\}\{\gamma(1+2\gamma z)^{-1/2}+\beta z'+\alpha r(1+z'^2)^{1/2}\}=z''\{(1+2\gamma z)^{1/2}-\beta r\}. \qquad (1)$$

For the purposes of the differential analyser this second-order, non-linear differential equation is best written as a set of four equations

$$r=\int\cos\psi\,ds, \quad z=\int\sin\psi\,ds, \quad V=\gamma\int dz/V,$$

$$\psi=\int\frac{(\beta\sin\psi+\alpha r+\gamma\cos\psi/V)ds}{V-\beta r}$$

where s is arc length and V is the ratio of velocity at a general point to the velocity of projection.

The results of the computation are presented in figures 1–3. In all this work the quantity of water flowing is taken to be 5 litres min^{-1}, and the surface tension of water is 72 dyn cm^{-1}. Figure 1 shows the shapes obtained when the water is projected horizontally, the parameters β and γ are the same for each curve ($\beta=\gamma=0{\cdot}12$) but α, which is proportional to the difference in air pressure inside and outside the bell, is varied. Curves are presented for $-0{\cdot}08\leqslant\alpha\leqslant+0{\cdot}08$ the limits of this range being determined by the fact that the pressure difference never exceeds 0·1 cm of water. Also shown in this figure is part of the parabola $\beta^2r^2-2\gamma z-1=0$, which is the 'limiting parabola' for the motion.

† Now at National Coal Board, Central Research Establishment, Isleworth, Middlesex.

Figure 2 shows three curves obtained when the water is projected downwards at an angle of 15° and three when it is projected downwards at 30° to the horizontal. The range of α is the same as for figure 1. Finally, figure 3 shows the effect of projection upward at 15° and 30° to the horizontal.

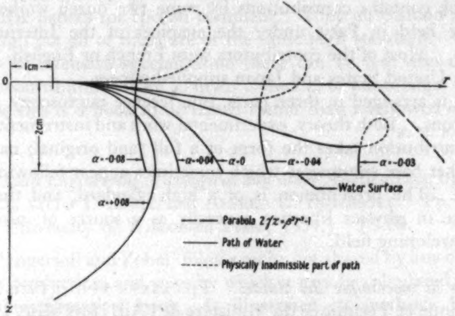


Figure 1. Horizontal projection from 0·14 cm slot.

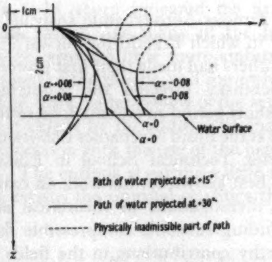


Figure 2. Effect of downward projection from 0·14 cm slot.

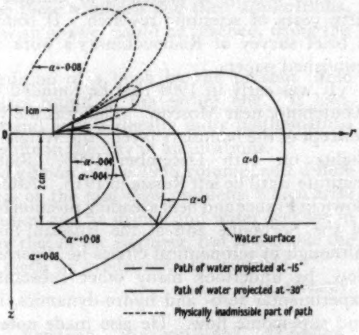


Figure 3. Effect of upward projection from 0·14 cm slot.

One writer (G. N. L.) is a National Academy of Sciences Research Fellow and he wishes to acknowledge a maintenance grant from the Foreign Operations Administration Scientific Research Project TA 01–101–3006 (OEEC 151). This work was sponsored, in part, by a University of California Research Grant.

Department of Engineering,
University of California,
Los Angeles 24,
California, U.S.A.

G. N. Lance.
E. C. Deland.

26th October 1954.

Lance, G. N., and Perry, R. L., 1953, *Proc. Phys. Soc.* B, **66**, 1067.

REVIEWS OF BOOKS

Contraste de Phase et Contraste par Interférences, by M. FRANÇON. Pp. 264. (Paris: Éditions de la Rèvue d'Optique Théorique et Instrumentale, 1952.) 2,200 *fr*. (2,500 *fr*. bound).

This book contains contributions by some two dozen workers in optics to a conference held in Paris under the auspices of the International Optical Commission. Most of the contributors were French or English, but in addition Sweden, the United States and Japan supplied papers.

The text is arranged in three parts, interference microscopy, phase contrast and applications. Both theory, experimental work and instruments are described and each contribution takes the form of a full (and original) paper as distinct from the rather bare summaries which sometimes appear following a conference of this kind. The presentation is of a high standard, and the work should take its place in physics libraries generally as a source of information in an important developing field.

H. H. HOPKINS.

Mémoires sur la mécanique des fluides. Pp. xxxv + 443. (Paris: Publications Scientifique et Technique du Ministère de l'Air, Hors série, 1954). 3,000 fr.

This volume is dedicated to Dimitri Riabouchinsky to commemorate his fifty years of scientific research. It comprises 39 contributions, preceded by a brief survey of Riabouchinsky's work and a chronological list of his many published papers.

It was early in 1904 that he founded the famous Aerodynamic Institute of Koutchino, near Moscow. This was the year in which Prandtl put forward the concept of the boundary layer; the Wright brothers had made their first powered flights on 17th December 1903. Riabouchinsky directed the Koutchino institute until he left Russia in 1918. After a short period in Denmark he settled down in France and held a leading position both at the Fluid Mechanics Laboratory of the Sorbonne and at the Russian Superior Technical School in France. Although in aeronautical circles he is perhaps best known for his work on cavity flow, he undertook many other substantial investigations in theoretical and experimental aero- and hydro-dynamics, including subsonic compressible flow and supersonic flow. He also made noteworthy contributions in the fields of ballistics, theoretical physics and mathematics.

His many and diverse researches are outlined in the introductory paper of the volume. This paper is in French and is followed by a very mediocre English version "traduced from french" (*sic*) with many misprints. Even the original is not free from error; for instance, it states, "La nouvelle imaginaire, $|j| = -1$, a conduit Dimitri Riabouchinsky à établir la théorie des fonctions des variables complexes $\phi + \epsilon\psi = f(x + \epsilon y)$, où ϵ^2 peut être une constante non seulement négative, mais aussi nulle ou positive". While this is presumably only a slip, it does not inspire confidence. One feels that this essay would have been far more satisfying if it had been less superficial and had endeavoured to relate some of Riabouchinsky's work to the state of knowledge at the time it was done, and to subsequent developments to which it gave rise. Without this it is scarcely possible to attain a right perspective.

As with Riabouchinsky's own researches, the papers which constitute the rest of the volume likewise embrace a wide range of subjects, covering various aspects of incompressible flow, subsonic compressible flow, supersonic flow and ballistics, together with certain other branches of fluid mechanics and topics from mathematics, classical mechanics and astrophysics. Most of the papers are in English or French, with three in German and one in Italian. From so wide a field and so many distinguished contributors it would not be proper to single out particular papers for special comment. They all contain something of interest, although not all of them are of the same high standard.

In these days of intensive specialization there can be few who could read and digest all the contributions from so many branches of knowledge. For most readers, therefore, this is a book to be tasted rather than swallowed or chewed.

R. C. PANKHURST.

Heat Conduction, with Engineering, Geological and other Applications, by LEONARD R. INGERSOLL, OTTO J. ZOBEL and ALFRED C. INGERSOLL. Pp. xiii + 325. (Madison: University of Wisconsin Press, 1954.) $ 5.00.

The pre-war 'Ingersoll and Zobel' filled a niche not shared by any competitor. It dealt with heat conduction, and ways by which it may be calculated, but it was by no means a mathematical book. It illustrated its methods by solving practical problems (with simple idealized conditions), but it was not an engineering textbook, which would have followed up these solutions and their applications. It was, in fact, a book which showed how an answer could be reached, using the simplest possible means.

In 1948, there appeared the first edition of a book by the present three authors, which had many of the features found in the earlier one. It had, however, grown a little more ambitious and gave a slightly more mathematical treatment of the theory, and somewhat more in the way of applications.

The work under review is the second edition of the 1948 volume, and differs from it chiefly by the addition of chapters on the heat-flow problems which arise in connection with the use of the heat pump, and on drying and water seepage in soil. The outlook is getting a little more that of the engineer, but as a whole the book retains the unusual qualities that made its predecessor so valuable.

J. H. AWBERY.

Electromagnetic Theory, by V. C. A. FERRARO. Pp. viii + 555. (London: Athlone Press, 1954.) 42*s*.

This book, written primarily to cover the requirements of an honours degree in mathematics, is naturally confined to the more familiar classical parts of the subject. Vector methods are used throughout, more advanced work in vector analysis and related topics being included in the first chapter. As befits a mathematical treatment, electrostatic, electromagnetic and gaussian units are used, practical units being defined in terms of them.

The first, and largest, part of the book covers the usual topics in electrostatics and magnetostatics. In particular, potential and force inside polarized and unpolarized media are discussed in more detail and more rigorously than is usual. The second part deals with boundary value problems governed by Laplace's equation which can be solved by means of conformal transformation or by the use of certain special functions (including spheroidal harmonics but excluding Bessel functions).

The third part, devoted to electrodynamics, includes such topics as electromagnetic potentials, self and mutual induction, Maxwell's equations, Poynting's theorem, simple a.c. circuit theory, induction in conducting media, reflection, transmission and refraction of electromagnetic waves and rectangular waveguides. There is also some work on retarded potentials and radiation from electric and magnetic oscillators and a simple account of the motion of electric charges in electric and magnetic fields (for small velocities).

The typography and layout of the book are very attractive; it is clearly written and pleasant to read. There is a good collection of problems at the end of each chapter and a fair number are worked out in full in the text.

A. N. GORDON.

Electronics, by A. T. STARR. Pp. vii + 395 (London: Pitman and Sons, Ltd., 1954.) 32*s*. 6*d*.

The main subject of this book is the circuits developed in conjunction with conventional electronic tubes, and hence its review belongs marginally in this journal because not all experimental physicists have reconciled themselves to their jools being mysterious boxes with knobs. On this subject the book contains tust about as much information as a physicist may aim at knowing. It is given in five chapters on valves, rectification, circuit theory, on amplifiers, oscillators and detectors, and on electronic applications. There is also an introductory chapter on physical fundamentals and a set of appendices on various aspects of basic mathematics.

In detail the book has, according to the author, been specifically written to cover the syllabus of the London University degree course on electronics. This course is additional to the telecommunication course and does apparently exclude everything connected with ultra high frequencies. The book obviously intends mainly to assist electronics students through the examinations, partly by frequent reference to past examination questions and partly by leaving out information useful to the general reader but not needed by the examinee. Thus, references to other books or to papers are completely absent. Also, for example, it seems somewhat arbitrary to find the cyclotron (stated to be suitable for electrons) explained but not the synchrotron, or, in the section on television camera tubes, to find various kinds of emitron discussed but not photo-conductive tubes. One is inclined to believe that this is because as a rule candidates are considered sufficiently knowledgeable if they know of one way of dealing with a particular problem.

The general reader, and perhaps the student as well, will also find himself somewhat discouraged by an occasional lack of precision. A particularly sad example of this may be quoted from the Introduction where the periodic table of elements is introduced thus: " Starting with a nucleus of atomic number 1, i.e., a proton, a neutral atom has only 1 electron. This is the hydrogen atom. The atom is normally in the 1s state, and we can write $H_1 = 1s^1$. It may be noted that the spin may be $+\frac{1}{2}$ or $-\frac{1}{2}$. It is interesting to note that two types of H_1 atoms exist, and they are called ortho- and para-hydrogen: their difference has effects at very low temperature ".

The main sections of the book, however, present on the whole an attractive and condensed, often (not always) very lucid survey of the fascinating field of valve and circuit engineering. It teaches the reader to calculate the properties of such circuits. Many physicists will find it worthwhile to read the book and even more to have it on their shelves for frequent reference. W. E.

A Selection of Graphs for use in Calculations of Compressible Airflow, by L. ROSENHEAD and others. Prepared on behalf of the Aeronautical Research Council by the Compressible Flow Tables Panel. Pp. x + 115. (Oxford: Clarendon Press, 1954.) 84*s*.

This book is a companion to the *Compressible Airflow Tables*, which was published in 1952, and the two volumes together complete the work of the Compressible Flow Tables Panel. Anyone who has occasion to undertake extensive calculations of the various quantities relating to high speed flow of a gas will inevitably consult these two books, and all that a reviewer need say is that the way has been made easy for them. The graphs in the present volume are a delight to the eye, and reflect an immense amount of loving care in their preparation. G. K. BATCHELOR.

Advances in Electronics, Vol. V, edited by L. Marton. Pp. xi + 420. (New York: Academic Press; London: Academic Books, 1953.) $9.50, 76*s*.

The editor and his collaborators deserve the highest credit for this fifth volume of the now well-established *Advances*. It contains eight articles, all 'front-line reports' in fields of the greatest actuality, all highly useful, while several of them are unusually brilliant.

The first article by R. Clark-Jones on the Performance of Detectors for Visible and Infra-red Radiation is a truly exemplary survey. It starts with an admirably clear and complete exposition of the mathematical theory of fluctuations, much of which is very recent, and has not yet found its way into textbooks. The enormous work which the author has done in not only listing but digesting 203 references, and by classifying and rating almost all detectors which are available at present deserves the gratitude of all workers in this field.

The second article by R. W. Hayward on Beta-Ray Spectrometers is a clear and simple introduction to the subject, with little mathematics. The third article by Ferd. E. Williams on Solid-State Luminescence is brilliant both as a physical introduction, and as a compendium of the most recent practical techniques, of which the production of transparent phosphors by Studer and Cusano is a particularly striking example. W. E. Danforth's article on Thorium Oxide in Electronics is an excellent and complete monograph. Weingartner and Kennedy give a review of Modern Vacuum Pumps in Electronics Manufacturing, of great interest to the practical valve engineer.

R. Q. Twiss, in an article on the Steady-State Theory of the Magnetron gives an impartial review of this difficult and controversial subject, and he gives also, for the first time in print, an all too short report of his own important work in his still unpublished Massachusetts Institute of Technology Thesis, in which he first marshalled really decisive arguments for the double stream theory.

The Review of Recent Work in Color Television, by Charles J. Hirsch is a truly admirable discussion, both theoretical and practical, of this most recent *tour de force* of the electronic art. The last article by J. S. Schaffner on Junction Transistor Applications is a masterpiece of efficient condensation and systematization.

As progress reports are now taking over more and more of the part which textbooks used to play in the past, this volume can be recommended not only as an almost indispensable reference book, but also as an introduction for students, in half a dozen different fields, from a common general ground right up to the front line.

D. GABOR.

Physical Properties of Solid Materials, by C. Zwikker. Pp. viii + 300. (London: Pergamon Press, 1954.) 60*s*.

In these days of highly specialized monographs it is refreshing to find a single volume which attempts to summarize the entire field of solid-state physics. As with most such small-scale maps, however, the details have often become blurred and are sometimes misleading.

The book deals with the fields of elasticity, plasticity, dislocations, ferromagnetism, semiconductors, electron theory of solids, surface effects and much else. Thus one finds accounts both of the use of sawdust in plastics and of Brillouin zones in wave-number space, and the variety of topics covered can be judged from the subject index, one section of which reads: Kelvin's solid—Kerr effect—Killers of fluorescence—Knudsen flow.

Clearly the book is intended as an introduction to the subject and the honours student may well profit from reading such a general survey before venturing into more detailed treatises. The author has not allowed himself to be confined by the conventional degree course treatment of his subjects and there is, for example, an interesting account of the Cauchy rariconstant theory of elastic moduli. Although this theory is now generally considered to be of only historic interest Professor Zwikker's lucid account of it does impress upon the reader the way in which the binding forces between the atoms in a crystal lattice are reflected in macroscopic elastic properties. On the other hand, Fermi–Dirac statistics are treated by analogy with the kinetic theory of chemical reactions and this is rather less convincing.

The book is well produced and is illustrated with many diagrams and half-tone plates. Although the reader may not always agree with the author's choice of subjects he should find the contents both interesting and stimulating.

H. KOLSKY.

Theoretical Physics—Mechanics of Particles, Rigid and Elastic Bodies, Fluids and Heat Flow, by F. W. Constant. Pp. xiv + 281. (Cambridge, Mass.: Addison-Wesley, *Principles of Physics* Series.) $6.50.

It is in many ways a point in favour of a textbook of mechanics that the subject matter and mode of presentation should be fairly conventional. The student meets with courses on mechanics at various levels of sophistication in his studies, and the courses and textbooks used should dovetail together readily. This new book by Professor F. W. Constant does in fact meet this requirement. It begins at a level suitable for the average student in an Honours course in Physics and follows the logical and well-tried line of development till it forms an introduction to post-graduate work.

The book is divided into three parts. Part I is a very clear and well-written set of three chapters on vector algebra, vector calculus and on dyadics, the latter being introduced for use in connection with principal axes of a rigid body, stresses and strains in elasticity, and rates of strain in viscous media. Part II, the main part of the book, contains the usual particle and rigid body dynamics including such topics as the linear oscillator, forced oscillation, coupled oscillations and normal coordinates, planetary motion and its inverse in α-particle scattering, Euler's equations for a rotating rigid body, and a brief introduction to Lagrange's and Hamilton's equations. In addition to particle and rigid body dynamics are chapters on elasticity and elastic waves, viscous media and fluid dynamics. Part III consists of a chapter on the mathematical formulation of heat flow, following logically after fluid dynamics with which of course it has much in common.

As mentioned before the general standard is about that which a student of good ability will attain during his undergraduate courses and the book will undoubtedly be used as a standard textbook by the more able student. The author states that the book is intended to bridge the gap between undergraduate and post-graduate courses and if by this is meant a thorough resumé and consolidation of undergraduate work, with a mathematical background suitable for the next step, then it is very successfully done. There is still, however, a considerable gap between this and, for example, *Classical Mechanics* by Goldstein which is a representative textbook on advanced mechanics.

The printing, illustrations, binding, etc., are all of high quality and there are plenty of problems to assist the process of consolidation. R. LATHAM.

Textbook of Physics, edited by R. KRONIG (translated from 3rd Dutch edition, 1951). Pp. xiv + 855. (London: Pergamon Press, 1954.) 70*s*.

This textbook is an English translation of a work by several eminent Dutch physicists. According to the publishers the original work has been well received in Holland but the reception which, in all probability, will be accorded here to the new version may not be more than a very mixed one. Even a casual perusal of the work makes one doubt the advisability of any authors or publishers making an attempt to give in one volume for degree students a really workable summary of physics. To the reviewer this book appears to fail in that it tries to combine things which are very elementary with others which are most advanced. Moreover, in order to find space, it seems as if the mathematical proofs have been shortened to such an extent that no typical student will be able to follow them.

The book also suffers from the fact that the notation used will be difficult for English students ; in particular the use of the letter O for surface area is very questionable here, although in the original no objection to it could be raised.

In a book of this nature it is surprising to find that there is no reference to any modern determination of the velocity of light or to Michelson's and later work on the metre. Since the measurement of spectral wavelengths has reached a degree of precision unequalled in any other branch of physics, and this is primarily due to the fact that Michelson did measure with great accuracy the wavelength of the red cadmium line with reference to the standard metre, such omissions are much to be deprecated.

Attention must also be called to some errors which should not have escaped the eyes of a team of experts. Only some can be mentioned. On p.100 we find a reference to a force–strain diagram when stress–strain is intended. On p.132 it is not pointed out that Poiseuille's equation is valid only for flow through a horizontal tube. The subject of non-Newtonian flow is not mentioned. On p. 309 Fig. 84 lines of force are shown cutting a metal surface at angles other than $\pi/2$. On p. 514 there appears a diagram showing the rise of a liquid in a capillary tube and the liquid therein is *flat*. In deriving a formula on p. 511 concerning falling drops, the writers appear to have ignored the work of Rayleigh and of Harkins and Brown.

On p. 405 there is an account of Michelson's stellar interferometer and here the errors are somewhat blatant. Rays fom a star must be parallel ; those shown do not satisfy this condition and the authors have not realized that Michelson only used a large telescope on account of the mechanical strength it possessed. The two mirrors which reflected light into the telescope actually covered only a small portion of the telescope objective. In the diagram they are shown covering nearly all the objective. This is an error common to many books. Michelson only covered a small portion of the aperture in order that the central ‘ blob ’ of the diffraction image should be large ; this was absolutely necessary since it was across this portion of the field that interference fringes were obtained, and then caused to disappear.

Finally, on p. 605 it is not made clear that the expression pv^γ only applies to a reversible adiabatic or isentropic change.

At the end of the book there is a somewhat comprehensive biographical note. This makes very interesting reading. The book is well illustrated and the production is good but the price of 70*s*. will put the book beyond the reach of nearly all students. C. J. SMITH.

Proceedings of the Toronto Meteorological Conference 1953. Pp. 294. (London : Royal Meteorological Society, 1954.) 30*s*.

In September 1953, the American Meteorological Society and the Royal Meteorological Society held a joint six-day meeting in Toronto, with the Canadian Branch of the Royal Meteorological Society acting as hosts. Nine symposia were held on subjects embracing in principle most of modern meteorology and the opening papers and discussions thereon have now been published as Proceedings in the format of the *Quarterly Journal of the Royal Meteorological Society*.

The papers vary in their aims. A few are in the nature of critical summaries, but not of a highly impersonal kind, while others report on original work, more or less complete or else in progress. The whole working face of the subject is thereby exposed and readers, or browsers, can thus obtain as good an idea of what meteorologists are currently ‘ up to ’ as in any volume published in recent years. To pick on a few meteorological subjects more or less at random and treated within these covers : the exploration of the high atmosphere, the distribution of infra-red cooling in the atmosphere, arctic meteorology, the global wind field and the jet-stream complex, numerical weather prediction, the correlation of climatic changes in different parts of the world, energy transfer from earth to atmosphere, and the growth of precipitation in clouds ; these give an idea of the range (though not the penetration) of the effort presently applied to the study of the atmosphere. The quality of the papers is inevitably uneven but the volume, like the occasion which prompted it, should become a not inconspicuous landmark in meteorological history. The Salisbury Press have made an attractive work of the printing.

P. A. S.

CONTENTS OF SECTION A

PAGE

Dr. R. L. F. Boyd and Dr. D. Morris. A Radio-Frequency Probe for the Mass-Spectrometric Analysis of Ion Concentrations 1

Dr. D. Morris. Molecular Ions in Discharges in the Inert Gases 11

Prof. S. Devons, Mr. G. Manning and Mr. D. St. P. Bunbury. Measurement of γ-Transition Lifetimes by Recoil Methods 18

Dr. A. H. Cooke and Dr. H. J. Duffus. The Magnetic Susceptibility of Potassium Manganicyanide 32

Dr. J. W. Weale. Alpha–Gamma Angular Correlations and Internal Conversion Measurements in ThCC″ and ThC′D 35

Research Notes:

Dr. D. F. Shaw. Scattering of 14·3 mev Neutrons by ^{4}He Nuclei . . . 43

Dr. A. Husain. Electron Scattering in Photographic Emulsions 45

Letters to the Editor:

Dr. J. A. Spiers. A Proposed Notation for Quantized Angular Momenta . 50

Mr. G. N. Harding. Absolute Calibration of Neutron Counters with Po–α–Be 52

Mr. E. Kroener. Dislocations and the Biot–Savart Law 53

Dr. B. Bleaney and Dr. W. Low. Nuclear Spins and Ratio of Magnetic Moments of Europium 151 and 153 55

Dr. B. Bleaney, Dr. G. S. Bogle, Dr. A. H. Cooke, Dr. R. J. Duffus, Miss M. C. M. O'Brien and Dr. K. W. H. Stevens. Paramagnetism of Caesium Titanium Alum 57

Reviews of Books 59

Contents of Section B 64

The Ionic Conductivity of Dilute Potassium Chloride Solutions at Centimetric Wavelengths

By V. I. LITTLE and V. SMITH
Bedford College, University of London

MS. received 21*st September* 1954 *and in final form* 10*th November* 1954

Abstract. A differential method is described for measuring the ionic conductivity of potassium chloride solutions at 3×10^9 c/s. The results reveal that a strong dispersion region exists at concentrations below 0·5 normal, which may be explained in terms of the perturbations by the applied field of a shell of water molecules surrounding the ion at a mean distance of 6 Å.

---

§ 1. Introduction

In the dipolar relaxation part of the frequency range, the variation with frequency (if any) of the ionic part of the conductivity of salt solutions is masked by the large apparent conductivity contributed by the dipoles of the solvent. The most detailed work published recently on the electrical properties of aqueous ionic solutions at 3×10^9 c/s and above has been that of Collie, Ritson and Hasted (1948) and Saxton and Lane (1952). These researches were most concerned, however, with the effect of the dissolved ions on the relaxation time of the solvent molecules. The smallest concentration of potassium chloride investigated was 0·5 normal, and it was convincingly demonstrated that the value of the ionic conductivity at this concentration, and for 1·0 normal, was almost certainly the same as the value at low frequency.

The present paper deals with an attempt to investigate the ionic conductivity effects of potassium chloride solutions in the concentration range 0·005 to 1·0 normal, in order to discover whether the relaxation process in the solvent could cause them to vary from their low-frequency value.

§ 2. Description and Theory of the Apparatus

The apparatus was designed to measure directly the difference in total conductivity between the solution under investigation and distilled water. As it was desirable to concentrate only upon conductivity changes, it was essential that the apparatus be insensitive to the changes in dielectric constant. It was decided that an H_{011} rectangular cavity with a small bore quartz tube in contact with the centre of one wall and parallel to the electric field, would be suitable. The changes in output of the cavity as distilled water and salt solution were passed alternately through the tube were interpreted in terms of conductivity changes.

A simplified sketch of the apparatus is shown in figure 1. The two liquids to be compared were passed through a temperature equalizer into a specially designed two-way tap which was driven through reduction gearing at 1 cycle in eight seconds by a constant speed motor. During one cycle of the tap, water flowed for three seconds, all flow ceased for one second, and then the solution flowed for three seconds followed by another waiting period of one second.

A chemical analysis of the last fraction of the outflowing liquids showed that under these circumstances, a complete substitution of the liquids in the tube was taking place. The temperature of the outflowing liquids was measured by means of a thermocouple, and another was used to monitor the temperature difference between the liquids just before they entered the tap.

A CV234 valve oscillator, amplitude modulated at 4330 c/s, and provided with two outputs, was used as a signal source tuned to a frequency of 3×10^9 c/s. Figure 2 gives a block diagram of the apparatus. The modulated signal was very loosely coupled to the H_{011} cavity, having passed through 10 dB of SAL. 1 type attenuating cable. An N.P.L. type cavity wavemeter was similarly coupled to the other oscillator output. The loosely coupled crystal detectors of both the H_{011} cavity and the wavemeter were connected through 100 dB of SAL. 2 type attenuating cable to two independent amplifiers, each of gain 100. The output of the wavemeter amplifier was passed through a circuit in which its phase could be varied continuously through 180° independently of its amplitude. It was then added electronically to the amplified H_{011} cavity signal, provision being made, by means of a variable resistance which could be switched in electrically, to vary the amplitude of this signal suddenly by a known small fraction.

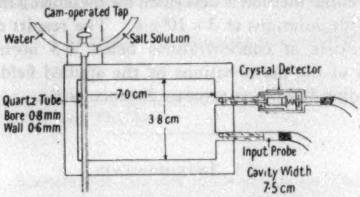

Cam-operated Tap
Water
Salt Solution
Crystal Detector
Quartz Tube
Bore 0·8 mm
Wall 0·6 mm
7·0 cm
3·8 cm
Input Probe
Cavity Width
7·5 cm

Figure 1. Cross section of H_{011} cavity for 10 cm waves.

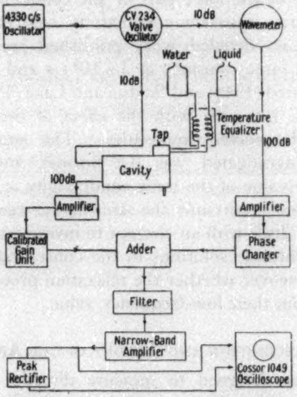
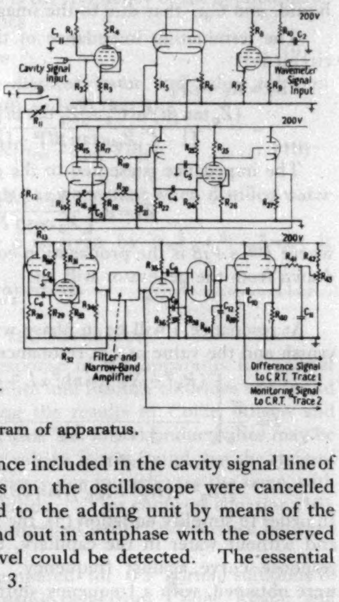

4330 c/s Oscillator
CV 234 Valve Oscillator
10 dB
Wavemeter
Water
Liquid
10 dB
Temperature Equalizer
Tap
100 dB
Cavity
100 dB
Amplifier
Amplifier
Calibrated Gain Unit
Adder
Phase Changer
Filter
Narrow-Band Amplifier
Peak Rectifier
Cossor 1049 Oscilloscope

Figure 2. Block diagram of apparatus.

The phasing unit was adjusted so that the wavemeter signal was in exact opposition to the cavity signal. The purpose of this arrangement was to increase the effective percentage modulation imposed on the 4330 c/s signal by the interchange of the liquids in the cavity tube. The change in level of the cavity signal when 0·005 normal solutions were under investigation was much less than 1%, but by using the above device the signal eventually fed to the high gain amplifiers could be converted to one effectively modulated to a depth of 50% or more. After passing through the adding unit, the signal was filtered in an inductance–capacity filter circuit, to remove most of the harmonic terms originating in the square law crystal detectors. Further amplification in a narrow-band amplifier of gain 10^4 was followed by peak rectification resulting in a signal fluctuating at the tap frequency. This signal was displayed on a Cossor model 1049 oscilloscope, together with the 4330 c/s signal, before final rectification. The latter signal was used for monitoring and tuning purposes.

When dealing with low concentrations, the ratio of the change in signal to the signal corresponding to distilled water in the cavity tube, could be obtained directly from the calibrated variable resistance included in the cavity signal line of the adding unit. The observed variations on the oscilloscope were cancelled by altering the amount of cavity signal fed to the adding unit by means of the calibrated control, and switching this in and out in antiphase with the observed change until no change in output signal level could be detected. The essential parts of this circuit will be found in figure 3.


200v
C_1
R_1
R_4
R_8
R_{10}
C_2
Cavity Signal input
R_2
R_3
R_5
R_6
R_7
R_9
Wavemeter Signal Input
R_{11}
200v
R_{15}
R_{17}
R_{19}
R_{23}
R_{25}
C_5
R_{12}
R_{14}
R_{16}
C_3
R_{18}
R_{20}
R_{21}
R_{22}
C_4
R_{24}
R_{26}
R_{27}
R_{13}
200v
R_{30}
R_{31}
C_7
C_6
R_{28}
R_{29}
R_{33}
R_{34}
R_{35}
C_8
C_9
R_{36}
R_{37}
R_{38}
R_{44}
C_{12}
R_{39}
C_{10}
R_{40}
C_{11}
R_{41}
R_{42}
R_{43}
R_{32}
Filter and Narrow-Band Amplifier
Difference Signal to C.R.T. Trace 1
Monitoring Signal to C.R.T. Trace 2

Resistances R
1, 2, 9, 10, 23, 24, 29, 30 = 750 kΩ
5, 6, 14, 15, 18, 22, 27, 28, 34, 35, 36 = 22 kΩ
12, 13, 20, 21, 32 = 150 kΩ ± 1%
4, 8, 25, 31, 42, 44 = 220 kΩ
3, 7, 26, 33 = 2·2 kΩ
39, 40 = 2 kΩ
41 = 2 MΩ
43 = 10 kΩ
16 = ½ MΩ variable
19 = 300 kΩ
11 = 50 kΩ and 10 kΩ calibrated.

Condensers C
1, 2, 4, 5, 6, 7 = 0·1 μF
8, 9, 10 = 0·01 μF
11 = 50 μF
3 = 0·0003 μF variable.

Figure 3. Circuit diagram of apparatus.

§ 3. CALCULATION OF THE APPARATUS RESPONSE TO CHANGE IN IONIC CONDUCTIVITY

Use is made of a transmission line analogue for the purpose of calculating the response of the H_{011} cavity to changes in the conductivity of the liquid in the quartz tube.

Figure 4 represents a short circuited transmission line of characteristic impedance Z_0, with a condenser of capacity C situated a short distance l from the short circuit. At a distance L from this condenser, the line is driven by a low impedance source of e.m.f. E connected in series to represent the loose coupling to the oscillator.

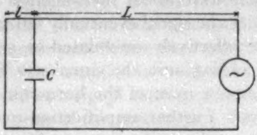


Figure 4. Equivalent line.

The capacity C may be written

$$C = C_1 - jC_2 \qquad \ldots\ldots(1)$$

where $C_1 = C_3 + C_0\epsilon'$ and $C_2 = C_0\epsilon''$, C_3 being the contribution due to the quartz tube, $C_0\epsilon'$ the contribution due to the real part of the dielectric constant of the liquid, and $C_0\epsilon''$ that due to the imaginary part.

The terminating impedance of the driven line at distance l from the short circuit is:

$$\begin{aligned} Z_T &= R_T + jX_T \\ &= \frac{(Z_0 \tan \beta l)^2 \omega\, C_2 + jZ_0 \tan \beta l\,[1 - C_1 Z_0 \omega \tan \beta l]}{[1 - C_1 Z_0 \omega \tan \beta l]^2 + [Z_0 \tan \beta l \omega C_2]^2}. \qquad \ldots\ldots(2) \end{aligned}$$

The impedance presented to the generator is given by

$$Z_I = Z_0 \left[\frac{(R_T + jX_T) \cosh PL + Z_0 \sinh PL}{Z_0 \cosh PL + (R_T + jX_T) \sinh PL} \right] \qquad \ldots\ldots(3)$$

where $P = \alpha + j\beta$ is the propagation constant of the line and the resulting current drawn from the generator will be

$$i = E/Z_1. \qquad \ldots\ldots(4)$$

At resonance, i will be in phase with E, so that the imaginary part of Z_I will vanish and the value of i at resonance is given by

$$i = \frac{E}{Z_0} \left[\frac{\begin{array}{r} (R_T{}^2 + X_T{}^2)(\sinh^2 \alpha L + \sin^2 \beta L) + Z_0{}^2(\cosh^2 \alpha L - \sin^2 \beta L) \\ + Z_0 (R_T \sinh 2\alpha L - X_T \sin 2\beta L) \end{array}}{\frac{1}{2}(R_T{}^2 + X_T{}^2 + Z_0{}^2) \sinh 2\alpha L + R_T Z_0 \cosh 2\alpha L} \right] \quad ..(5)$$

with the resonance condition that

$$\tfrac{1}{2}[Z_0{}^2 - (R_T{}^2 + X_T{}^2)] \sin 2\beta L + Z_0 X_T [\cosh^2 \alpha L - \sin^2 \beta L] = 0.$$

In order to simplify equation (5), the Q factor of the cavity was determined with and without water in the capillary, by the method of plotting the square law response curve against frequency. The values 1500 and 2000 respectively were obtained, with a frequency shift of less than 1 in 3000 Mc/s. No further frequency shift could be detected when normal potassium chloride solution was substituted for distilled water. The large value of the Q factor even with water or saline present in the tube implied that R_T was very small compared with Z_0. The very slight change in tuning was taken to imply that X_T also was small compared with Z_0, and that βL was very nearly π.

Using these experimental conclusions, equation (5) may be simplified considerably to

$$i = \frac{E}{Z_0} \left[\frac{Z_0 \cosh^2 \alpha L}{\frac{1}{2} Z_0{}^2 \sinh 2\alpha L + R_T Z_0 \cosh 2\alpha L} \right]. \qquad \ldots\ldots(6)$$

Substituting for R_T from the real part of equation (2), and neglecting terms containing $\tan \beta l$ in the denominator, we obtain

$$i = \frac{E}{Z_0}\left[\frac{\cosh^2 \alpha L}{\frac{1}{2} Z_0 \sinh 2\alpha L + \cosh 2\alpha L\,(Z_0 \tan \beta l)^2 \omega C_2}\right] = \frac{1}{A + B\epsilon''} \qquad \ldots(7)$$

where A and B are constants, and equation (1) has been used to substitute for C_2.

In the cavity, i is directly proportional to the maximum tangential magnetic field stimulating the detector, and the e.m.f. injected into the detector circuit will be proportional to i.

The imaginary part of the dielectric constant of saline solution may be considered to be made up of two parts:

$$\epsilon_s'' = (2\sigma_{DS} + 2\sigma_{IS})/\nu \qquad \ldots\ldots(8)$$

where σ_{DS} represents the solution conductivity due to dipolar relaxation and σ_{IS} the high frequency ionic conductivity. ν is the frequency of the applied field.

For distilled water the expression becomes

$$\epsilon_W'' = 2\sigma_{DW}/\nu. \qquad \ldots\ldots(9)$$

Using the suffixes W and S to refer to distilled water and saline solution respectively, it follows from equations (7), (8) and (9) that

$$\frac{i_W - i_S}{i_S} = \frac{B(\epsilon_S'' - \epsilon_W'')}{A + B\epsilon_W''} = K[\sigma_{DS} - \sigma_{DW} + \sigma_{IS}] \qquad \ldots\ldots(10)$$

where $K = 2B/[A + B\epsilon_W'']\nu$ and can be regarded as a constant when distilled water is used as the standard.

The experimentally measured quantity δ is given by $(i_W^2 - i_S^2)/i_W^2 = \delta$; it therefore follows that

$$(i_W - i_S)/i_S = (1-\delta)^{-1/2} - 1. \qquad \ldots\ldots(11)$$

§ 4. CALIBRATION OF THE APPARATUS

For the purposes of calibration, the response of the apparatus for normal and half normal solutions of potassium, sodium and lithium chlorides was found at temperatures from 18°C to 25°C. Using the results of Collie, Ritson and Hasted (1948) for these concentrations, i.e. that the ionic conductivities may be assumed to be equal to their low-frequency values, it was found that the ratio of $(i_W - i_S)/i_S$ to the corresponding d.c. conductivity was in each case the same, with a maximum deviation of 2%, for lithium chloride. This confirms equation (10), in which one would expect σ_{IS} to be large compared with the difference between σ_{DS} and σ_{DW}.

In table 1, σ_{IS} and $\sigma_{DW} - \sigma_{DS}$ are compared for 0·5 normal solutions of potassium and lithium chloride at 25°C, using data from Collie, Ritson and Hasted (1948). The value of ϵ_{DW}'' at 25°C, has been taken as 11·3.

Table 1

	$\frac{2}{3}\sigma_{IS} \times 10^{-9}$	$\frac{2}{3}(\sigma_{DW} - \sigma_{DS}) \times 10^{-9}$
0·5 n KCl	35·2	0·8
0·5 n LiCl	23·9	0·6

Having thus determined K in equation (10), the values of $(i_W - i_S)/i_S$ could be interpreted directly in terms of the ionic conductivity of saline solutions.

§ 5. Experimental Results

The experimental results for potassium chloride solutions are expressed graphically in figure 5, where the ratio of the high-frequency conductivity to the low-frequency value is plotted against concentration in gramme equivalents per litre. The low-frequency conductivity data were obtained from the international critical tables and from conductivity bridge measurements. Sodium and lithium chlorides were also investigated, but these salts did not show such marked effects as potassium chloride. Lithium chloride showed no effect which could be differentiated from the experimental error, but sodium chloride exhibited the same behaviour as potassium chloride at much lower concentrations. The results for sodium chloride are given in table 2.

Table 2

Concentration (g equiv litre)	1	0·5	0·2	0·1	0·0128	0·01	0·0085	0·005
Ratio, 18°c	1	1	1	1	1·29	1·29	1·30	1·35
Ratio, 25°c	1	1	1	1	1·29	1·34	1·36	1·44

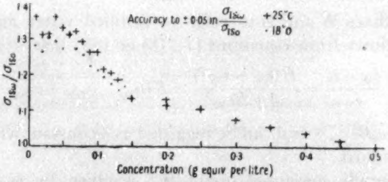


Figure 5. Variation of ratio of high-frequency to low-frequency ionic conductivity with concentration.

The experimental error associated with these results is about ±0·05 units on the vertical scale of figure 5 for concentrations below 0·1 normal, but considerably less than this at the higher concentrations. Both sodium and potassium chlorides exhibit a marked increase in conductivity over their low-frequency values in dilute solutions, amounting to about 30%. Such an increase at these concentrations cannot be due to an ion atmosphere relaxation effect of the kind described by Falkenhagen (1934), because the latter effect in 0·01 normal solution would only amount to a 1 or 2% increase. It seems very probable that the result is due to a solvent atmosphere effect, particularly since the conductivity increases as the concentration decreases. A tentative theory based on the interaction between an ion and its solvent atmosphere is discussed in the following section.

§ 6. Interpretation of Results

The large increase in the high-frequency ionic conductivity of dilute potassium chloride solutions, may be interpreted semi-quantitatively in terms of the interaction of an ion with its surrounding sheath of water molecules. Bernal and Fowler (1933) have pointed out that the intense electrostatic field due to an ion will have a disrupting effect on the structure of the water surrounding it. They postulate that within a certain critical distance from the ion, the water dipoles will be co-ordinated by the ion, while beyond this distance they will be orientated by their mutual dipole attractions. Debye (1929) has calculated that within a sphere of radius less than 12 Å about a monovalent positive ion, the dielectric constant of water is very much modified. In order to interpret the present results, however, it is necessary to make some assumptions concerning the electrostatic field due to an ion at distances less than 10 Å, and the number of water dipoles in successive layers about the ion which are strongly orientated by it. Pople (1951) has calculated the number of water molecules associated with any one molecule using the experimental results of Morgan and Warren (1938) on the x-ray analysis of water. He finds that a good fit between theory and practice occurs if the first three shells of neighbours surrounding a molecule contain 4, 11 and 22 molecules. A perfect tetrahedral structure would imply 4, 12 and 36 for the numbers in these shells.

We shall assume that in the case of an ion, the radially disposed water molecules surrounding it are packed into shells containing 4, 12 and between 22 and 36 molecules. In addition it is assumed that the inner two shells are packed down closer to each other than they are in the normal water structure, but that the third shell remains at a distance of about 6 Å from the ion which is its position with respect to a central molecule in pure water. The water molecules forming shells one to three are considered to be strongly orientated by the field due to the ion, but beyond shell three the normal tetrahedral structure of water starts to reassert itself. The object of these restrictions is to produce a shell of radially disposed dipoles, which are fairly free from the influence of the dipoles within it.

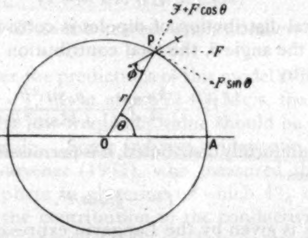


Figure 6.

When calculating the electrostatic field at the third shell due to the central ion, it is not possible to use an expression for the dielectric constant (derived in terms of a continuous medium in a high state of saturation), because the medium within the shell consists simply of one ion and sixteen water molecules. It is assumed therefore that the field due to the central ion is fully effective through the gaps in the screen formed by the dipoles of the inner shells. The majority of the dipoles of shell three then move to positions where this field is a maximum, leaving gaps which can be filled by dipoles from regions beyond shell three. In this way the increase in density due to the presence of ions can also be explained, in spite of the fact that shell three is assumed to remain at a fixed radius from the central ion.

In figure 6, P represents a dipole in the third shell which has been rotated out of the radial direction at the point (r, θ) by the application of an external field F. If a monovalent positive ion is assumed to exist at O, and the applied field is in the direction OA, it follows that the dipole at P rotates through an angle ϕ given by

$$\tan\phi = \frac{F\sin\theta}{\mathscr{F} + F\cos\theta} \qquad \text{......(12)}$$

where

$$\mathscr{F} = \boldsymbol{e}/r^2, \qquad \text{......(13)}$$

$\boldsymbol{e}$ being the electronic charge, and $\mathscr{F}$ the radial field due to the ion at a distance r cm.

If $F\cos\theta \ll \mathscr{F}$, it follows that

$$\tan\phi = k\sin\theta \quad \text{where} \quad k = F/\mathscr{F} \qquad \text{......(14)}$$

and

$$\cos\phi = (1 + k^2\sin^2\theta)^{-1/2}, \quad \sin\phi = k\sin\theta(1 + k^2\sin^2\theta)^{-1/2}. \qquad \text{......(15)}$$

The contribution to the field at O along OA by the dipole at P is given by

$$\delta f = \frac{\mu}{r^3}[2\cos\theta\cos\phi - \sin\theta\sin\phi]$$

where μ is the effective dipole moment of the water molecule. Substituting from equation (15), this becomes

$$\delta f = \frac{\mu}{r^3}\left[\frac{2\cos\theta - k\sin^2\theta}{(1 + k^2\sin^2\theta)^{1/2}}\right].$$

If a symmetrical distribution of dipoles is considered, in which a number n_θ is associated with the angle θ, the total contribution of the shell to the field at O along OA is given by

$$f = -\frac{k\mu}{r^3}\sum\frac{n_\theta\sin^2\theta}{(1 + k^2\sin^2\theta)^{1/2}}. \qquad \text{......(16)}$$

If n dipoles are uniformly distributed, it is permissible to integrate equation (16), which then becomes

$$f = -\tfrac{2}{3}k\mu n/r^3. \qquad \text{......(17)}$$

The value of μ is given by the Langevin expression

$$\mu = \mu_0\left[\coth\frac{\mu_0\mathscr{F}}{\boldsymbol{k}T} - \frac{\boldsymbol{k}T}{\mu_0\mathscr{F}}\right] \qquad \text{......(18)}$$

where μ_0 is the true dipole moment of water $= 1{\cdot}87 \times 10^{-18}$ e.s.u., $\boldsymbol{k}$ is Boltzmann's constant $= 1{\cdot}37 \times 10^{-16}$, T is the absolute temperature $= 298°$.

Using the value $4{\cdot}8 \times 10^{-10}$ e.s.u. for the value of $\boldsymbol{e}$ in equation (13), and putting $r = 6$ Å, equation (17) becomes on substitution of equations (13), (14) and (18)

$$f = -0{\cdot}035nF. \qquad \text{......(19)}$$

The resultant field acting upon the ion is thus

$$F + f = F[1 - 0{\cdot}035n]. \qquad \text{......(20)}$$

The fact that a positive ion will move in the direction of the applied field sets the upper limit to n at the value $1/0{\cdot}035$ i.e. $n_{\max} = 28$ molecules.

It must be noted that this number represents the maximum number of molecules in the third shell which are fully available to contribute a maximum reaction to the field at the ion; it does not necessarily represent the total number of molecules in the shell.

At sufficiently high frequencies, the molecules of the third shell will lag behind the field as they swing about their equilibrium position, and at the same time the amplitudes of their swings will diminish. Equation (20) may be rewritten in a general form as follows:

$$R_\omega = F\left[1 - \frac{0{\cdot}035n}{1+\omega^2\tau^2}\right] \qquad \ldots\ldots(21)$$

where R_ω now represents the resultant field parallel to and in phase with the applied field F, τ is the relaxation time of the molecules of the shell, and ω the angular frequency of the applied field. Equation (10) also expresses the variation of conductivity with frequency according to this model. The ratio of the high-frequency conductivity to the low-frequency conductivity is given by

$$\frac{R_\omega}{R_0} = \frac{\sigma}{\sigma_{\mathrm{IS0}}} = \frac{1+\omega^2\tau^2 - 0{\cdot}035n}{(1+\omega^2\tau^2)(1-0{\cdot}035n)}. \qquad \ldots\ldots(22)$$

A suitable value of τ is chosen corresponding to the relaxation wavelength of concentrated solutions of potassium chloride as measured by Collie, Ritson and Hasted (1948), namely 1·4 cm for 2-normal solutions. In the present work, ω was $2\pi \times 3 \times 10^9$ whence $\omega\tau$ becomes 0·14. The observed ratio for $\sigma_{\mathrm{IS}\omega}/\sigma_{\mathrm{IS0}}$ in the present work was 1·3. Substituting these values into equation (22) we have

$$1{\cdot}3 = \frac{1+(0{\cdot}14)^2 - 0{\cdot}035n}{(1+(0{\cdot}14)^2)(1-0{\cdot}035n)}$$

whence $0{\cdot}035n = 0{\cdot}94$ and $n = 27$ to the nearest whole number, in good agreement with the theory.

It is interesting to consider the predictions of this model at higher frequencies. At frequencies such that $\omega\tau = 1$, i.e. at about 21 400 Mc/s, the ratio of the high-frequency conductivity to the low-frequency value should be about 5 for dilute solutions of potassium chloride. Some indirect evidence in support of this is contained in a paper by Gärtener (1931), who measured the conductivity of solutions of magnesium sulphate in glycerine to which 4% of water had been added. After allowing for the contribution to the conductivity due to dipolar relaxation in the glycerine, the observed ionic conduction was some five times the low-frequency value in concentrations of 0·1 molar. Bearing in mind the fact that the magnesium ion is divalent, the radius of the equivalent shell is extended to 8·5 Å, which, taken in conjunction with the larger size of the glycerine molecule, could well produce effects similar to the results observed here for aqueous potassium chloride solutions.

Equation (21) may be written in a more generalized form as

$$R_\omega = F\left[1 - \frac{\psi(\mu_0 e'/a)}{1+\omega^2\tau^2}\right] \qquad \ldots\ldots(23)$$

where $\psi(\mu_0 e'/a)$ is some function of the expression in parentheses, e' being the charge on the ion, and a the ionic radius. If this equation should represent the field acting upon an ion, the following statements may be made: (i) The low frequency conductivity should decrease as e'/a increases. (ii) At high frequencies an anomalous dispersion of conductivity should be observed provided $\omega\tau$ approaches unity. At high concentrations τ is likely to diminish due to the proximity of positive and negative ions, and this would cause the dispersion effect to be moved bodily to higher frequencies. The observed decrease in the dispersion effect for potassium chloride as the concentration was increased could be explained in this way. (iii) At very high concentrations, the proximity of ions of like charge would be a 'shell-breaking' effect leading to an increase in conductivity. This increase, however, would be offset by the dissymmetry and electrophoretic effects described in the Debye–Hückel (1923) theory of electrolytes. In addition, ionic association of the type postulated by Bjerrum (1926) would cause a further decrease in conductivity. The combined action of these effects could lead to the observed linear decrease in the conductivity of concentrated solutions of strong electrolytes at low frequencies. Under these circumstances, the only high-frequency dispersion effect would be that described by Falkenhagen (1934). Note that in the above theory τ refers to the relaxation time of a water molecule in the presence of the strong field due to an ion. The effect of the field due to the ion superposed upon the mutual dipole forces would be to shorten the relaxation time of a dipole in a shell, because such dipoles would have a preferred orientation toward the ion. A neighbouring ion of opposite sign would tend to increase the orienting field, and thus reduce the relaxation time even further.

A more exact theory of this effect must be able to account for the large difference in concentration at which the effect is observed in potassium and sodium chloride solutions.

Further work at a frequency of 3×10^9 c/s on ions of higher valency is in course of preparation, and will be published later.

ACKNOWLEDGMENTS

The authors wish to thank the members of the physics and chemistry departments of Bedford College for their helpful criticisms during the course of this work. In particular our thanks are due to Mr. F. Grimes of the Physics Department for the design and manufacture of a tap controlling the flow of liquids in the apparatus. Mrs. V. Smith was working on behalf of the British Empire Cancer Campaign and thanks them for their continued help during the course of this investigation.

Due acknowledgment is also made to the Chief Superintendent, Radar Research and Development Establishment, for the loan of an N.P.L. type cavity wavemeter.

REFERENCES

BERNAL, J. D., and FOWLER, R. H., 1933, *J. Chem. Phys.*, **1**, 515.
BJERRUM, N., 1926 a, *K. Danske Vidensk. Selsk. Mat. Fys. Medd.*, **7**, 9; 1926 b, *Ergebn. exakt. Naturw.*, **6**, 125.
COLLIE, C. H., RITSON, D. M., and HASTED, J. B., 1948, *J. Chem. Phys.*, **16**, 1.
DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalog Co.).
DEBYE, P., and HÜCKEL, E., 1923, *Phys. Z.*, **24**, 185, 305.
FALKENHAGEN, H., 1934, *Electrolytes* (Oxford : University Press).
GÄRTENER, H., 1931, *Phys. Z.*, **32**, 919.
MORGAN, J., and WARREN, B. E., 1938, *J. Chem. Phys.*, **6**, 666.
POPLE, J. A., 1951, *Proc. Roy. Soc.* A, **205**, 163.
SAXTON, J. A., and LANE, J. A., 1952, *Proc. Roy. Soc.* A, **214**, 531.

Thermal Expansion of Potassium Chlorate

By M. A. LONAPPAN
Department of Physics, University of Madras, Guindy, Madras, 25

Communicated by G. N. Ramachandran ; *MS. received 24th August* 1954

Abstract. The principal coefficients of expansion and the orientation of the ellipsoid of expansion have been determined for potassium chlorate for the ranges 30 to 90, 90 to 150 and 150 to 200°c. An x-ray method was used and the expansion coefficient in the directions normal to the various planes in the zones [100], [010] and [001] were determined by recording the zero layer of the rotation photographs taken about the three axes in a Unicam high temperature camera. The values of α_{22} and α_{33} are nearly equal (~ 30 to 40×10^{-6}) whereas the value of α_{11} ($\sim 130 \times 10^{-6}$) is much higher than both. The direction of the greatest expansion coefficient is found to be almost perpendicular to the plane of the oxygen atoms according to Zachariasen's structure. The expansion coefficients along *a*, *b* and *c* axes and the rate of change of the monoclinic angle with temperature are also given, using which the lattice dimensions of the crystal at any temperature can be determined.

§ 1. Introduction

In this paper are reported the results of an experimental study of the thermal expansion of potassium chlorate. This investigation arose as an off-shoot of an attempt to find out the modifications in the structure, if any, which take place at the transition temperature of 250°c. It has been observed that if an untwinned crystal is heated above this temperature and then cooled it becomes multiply twinned (Madan 1886, Rayleigh 1888, 1889, Sirkar 1930).

Potassium chlorate crystallizes in the class $2/m$ (C_{2h}) and belongs to the space-group $P2_1/m$ with two molecules per unit cell (Zachariasen 1929). Since the crystal is monoclinic, the ellipsoid describing its expansion tensor is triaxial. One of the axes, namely α_{22} axis, coincides with the *b* axis of the crystal, while the other two axes lie in the *ac* plane and in order to describe completely the ellipsoid it is necessary to find also the angle made by α_{11} axis with the *c* or *a* axis of the crystal. A complete determination of this type using Fizeau's method for the different sections of the crystal has been made on cane sugar (monoclinic) and copper sulphate (triclinic) by Sreedhar (1952, 1953). X-ray methods have, however, been used only for crystals of monoclinic or higher symmetry. Ubbelohde and co-workers (1939, 1946) have determined the thermal expansion of oxalic acid (monoclinic) and Rochelle salt (orthorhombic). Rosemary Shaw (1953) has reported the thermal expansion of afwillite, a monoclinic crystal determined by x-ray methods.†

---

† The author came to know of these studies by the x-ray method only after the investigations reported in this paper were completed and a literature survey was made before writing up the paper. The methods followed here are similar to those adopted by the earlier workers, except for the experimental arrangement and the greater accuracy.

The present study of thermal expansion of potassium chlorate has been made with a good degree of accuracy for three ranges of temperature, namely 30 to 90, 90 to 150 and 150 to 200°c. This was done making use of a Unicam high temperature camera and by studying the variations in the spacings of the reflections in the *a*, *b* and *c* zones.

§2. EXPERIMENTAL METHOD

Rotation photographs were taken about the *a*, *b* and *c* axes and the zero-layer reflections were recorded using a 19 cm camera (Unicam). The spacings of the high-angle reflections could be accurately measured with this camera, using the fiduciary marks at the high angle end and the changes in spacings with increase of temperature could thus be accurately known. In order to avoid the overlap of reflections having nearly the same θ-values the 'tilted crystal method' was used, i.e. tilting the crystallographic axis slightly away from the rotation axis by 2 to 3°. In order to index the high-angle reflections a Weissenberg photograph of the zero layer was also taken using the same crystal. In this way the thermal expansion perpendicular to a number of crystallographic planes in the zone concerned could be determined. These were used for calculating the principal expansion coefficients and the orientation of the principal axes in the particular zone.

Three zones were studied by taking rotation photographs about the *a*, *b* and *c* axes. Although potassium chlorate normally crystallizes in a tabular form with large *c*-faces, a few needle-shaped crystals were found having *a* and *c* axes as needle axis (Groth 1910). These were used for taking the *a* and *c* rotation photographs. A few crystals also occurred with [110] as the needle axis and these were used to obtain the *b* rotation photographs. Such a crystal was fixed to a glass fibre at a suitable angle using fire-clay, the latter being attached to the eccentrics of the high temperature camera through a brass stud. The fire-clay proved to be unsatisfactory above 200°c.

§3. METHOD OF CALCULATION

The thermal expansion coefficient α for a particular reflection is determined using the relation

$$\alpha = -\frac{1}{\Delta t}\frac{\Delta(2\sin\theta)}{2\sin\theta}$$

where Δt is the range of temperature. The distance of the spot concerned from the fiducial 85° mark on the film was measured and $2\sin\theta$ was calculated from it. The magnitude of $\Delta(2\sin\theta)$ was obtained as the difference between the measured values of the two temperatures. Only reflections having a value greater than 60° were used for the calculation of thermal expansion. In the case of *a* and *c* rotation photographs five such high-angle reflections were measured.

From the indices of a plane one can calculate the inclination ϕ of its normal to a chosen direction in the zone. In the case of *a* and *c* axis rotation photographs, the *b* axis is taken as the direction of reference. Then the relation between α_ϕ and ϕ takes the following form:

For the [100] zone $\alpha_\phi = \alpha_{22} + \sin^2\phi(\alpha_{c^*} - \alpha_{22})$ ......(1)

For the [001] zone $\alpha_{\phi'} = \alpha_{22} + \sin^2\phi'(\alpha_{a^*} - \alpha_{22})$. ......(2)

We have five equations of type (1) and five equations of type (2) from which α_{22}, α_{a^*} and α_{c^*} are to be determined. This was done by the method of normal equations in one step using all the ten equations and considering α as a function of both ϕ and ϕ'. Suitable weights were given to the different α values according to their nearness to 90°.

In the case of b axis rotation photographs eight high-angle reflections could be measured. Here, if ξ is the angle made by the normal to the plane with a chosen direction (the c^* axis), then α_ξ may be put in the form:

$$\alpha_\xi = A + B\cos 2\xi + C\sin 2\xi \qquad \ldots\ldots(3)$$

where A, B and C are functions of α_{11}, α_{33} and ψ, the inclination of α_{11} axis to the c^* direction. The relations are given by (Wooster 1949)†

$$\left.\begin{aligned} A &= \tfrac{1}{2}(\alpha_{11}+\alpha_{33}) \\ B &= \tfrac{1}{2}(\alpha_{11}-\alpha_{33})\cos 2\psi \\ C &= \tfrac{1}{2}(\alpha_{11}-\alpha_{33})\sin 2\psi \end{aligned}\right\}. \qquad \ldots\ldots(4)$$

In addition to the eight observed values, the values for α_{a^*} and α_{c^*} obtained from the [001] and [100] zones were also fed in (with a weight of five each), so that there were ten equations of type (3). These were solved by the method of normal equations to give the most probable values for A, B and C. The values of α_{11} and α_{33} and ψ were obtained from them by solving the equations (4).

The probable errors were calculated using standard methods (Whittakar and Robinson 1944). In the case of [010] zone alone since the quantities finally required, namely α_{11}, α_{33} and ψ, are not linearly related to α_ξ it was necessary to use the three functions A, B and C which are linearly related as in equation (3). A suitable mathematical technique was developed to calculate the probable errors of α_{11}, α_{33} and ψ.

§4. RESULTS OF THE INVESTIGATION

The measured values of the expansion coefficients perpendicular to various planes in the three zones studied are given in tables 1 and 2. The values α_{22}, α_{a^*} and α_{c^*} calculated from a and c axis rotation photographs are given in table 3.

Table 1. Measured Thermal Expansion Coefficients of the Reflections in the [100] and [001] Zones

(1)	(2)	$\sin^2\phi$	α_ϕ ($\times 10^6$) for 30–90°c	90–150°c	150–200°c
072	13° 27′	0·0541	44·0	37·9	44·6
047	55° 41′	0·6821	83·7	79·9	99·4
028	73° 22′	0·9180	102·5	99·6	111·8
008	90° 00′	1·0000	108·8	108·6	109·0
055	39° 56′	0·4120	66·0	65·4	67·4
170	10° 20′	0·0322	39·9	34·9	34·8
530	64° 49′	0·8188	80·8	73·4	84·3
440	51° 56′	0·6198	71·0	66·6	73·4
510	81° 06′	0·9761	89·2	77·4	96·1
350	37° 26′	0·3693	54·8	50·1	59·7

(1) Indices of reflections, (2) inclination ϕ with b^* axis.

The final values of α_{11}, α_{22}, α_{33} and ψ together with their probable error are given in table 4. In order to bring out the agreement between the observed values and those calculated from the constants of the ellipsoid of thermal expansion, figures 1 and 2 are drawn. In figure 1 α is plotted against $\sin^2\phi$ or $\sin^2\phi'$ as the case may be. The observed values should lie on two straight lines and they are found to lie very close to the theoretical values. Figure 2 gives a polar diagram of α_ξ as well as that of $1/\sqrt{\alpha_\xi}$. In each case the continuous curves represent the variations to be expected theoretically. Here again the observed values are seen to lie close to the theoretical values.

---

† The sign convention adopted here is different from that of Wooster.

Table 2. Measured Thermal Expansion Coefficients of the Reflections in [010] Zone

		α_ξ ($\times 10^6$) for		
(1)	(2)	30–90°c	90–150°c	150–200°c
$60\bar{3}$	89° 31′	53·8	53·9	61·9
207	19° 45′	133·6	122·7	146·2
502	57° 34′	113·3	105·5	124·3
$\bar{2}09$	160° 13′	78·6	62·6	74·4
404	43° 30′	125·6	114·7	—
$40\bar{8}$	136° 02′	40·0	30·7	—
008	0	111·4	102·2	108·9
$50\bar{6}$	115° 37′	29·7	30·0	26·2

(1) Indices of the reflections, (2) inclination ξ with c^* axis.

Table 3. Values of α_{22}, α_{a^*} and α_{c^*} calculated from the Data in Table 1

Temp. range (°c)	30–90	90–150	150–200
$\alpha_{22} \times 10^6$	39·0	33·8	36·8
$\alpha_{a^*} \times 10^6$	90·0	82·7	95·5
$\alpha_{c^*} \times 10^6$	107·2	104·9	120·1

Table 4. Elements of the Expansion Ellipsoid of $KClO_3$

α_{11}, α_{22} and α_{33} are the principal expansion coefficients and ψ is the angle between α_{11} direction and the normal to the c plane

Temp. range (°c)	30–90	90–150	150–200
ψ	$29·6 \pm 0·5°$	$29·2 \pm 0·7°$	$29·6 \pm 0·6°$
$\alpha_{11} \times 10^6$	$134·2 \pm 1·3$	$126·6 \pm 1·9$	$148·8 \pm 1·9$
$\alpha_{22} \times 10^6$	$39·0 \pm 0·6$	$33·8 \pm 0·9$	$36·8 \pm 2·1$
$\alpha_{33} \times 10^6$	$29·7 \pm 2·1$	$26·4 \pm 3·0$	$27·6 \pm 3·3$

Table 5. Rate of Variation of Unit Cell Dimensions with Temperature

Temp. range (°c)	30–90	90–150	150–200
$\alpha_a \times 10^6$ (deg^{-1})	55·1	50·2	57·2
$\alpha_b \times 10^6$ (deg^{-1})	39·0	33·8	36·8
$\alpha_c \times 10^6$ (deg^{-1})	74·4	69·8	79·3
$d\beta/dt \times 10^6$ (rad deg^{-1})	96·5	92·4	112·0

It is noticed that all the thermal expansion coefficients decrease from the first range, viz. 30 to 90°c, to the second range, viz. 90 to 150°c, and they increase again from the second to the third range, viz. 150 to 200°c. A careful check was made to see whether this was due to any error in temperature measurement. A recalibration of the thermocouple showed that the readings could not be in error by more than 1°c. The peculiar variation of the expansion coefficient with temperature therefore appears to be real. This is particularly so since the results obtained from the different zones of the crystal agree with one another for the same range of temperature. However, it may be mentioned that neither the orientation of the ellipsoid nor its general appearance is appreciably altered in this range of temperatures.

It is interesting that α_{22} and α_{33} are found to be nearly equal whereas the value of α_{11} is much higher than both. This can be attributed to the occurrence of groups of three oxygen atoms in the ClO_3 ion. In calcite also similar O_3 groups are found and all the O_3 planes are parallel. The resemblance between the structures of calcite ($CaCO_3$) and potassium chlorate has already been pointed out by Zachariasen (1929). The similarity between the structures of sodium nitrate ($NaNO_3$) and potassium chlorate is perhaps even closer, since both contain only alkali metals and univalent ions of the type BO_3. In fact, the values of the coefficient of thermal expansion of $KClO_3$ and $NaNO_3$ are of the same order of magnitude. In the structure of $KClO_3$ as determined by Zachariasen the normal to the plane of the oxygen atoms lines in the (010) plane and makes an angle of 53·9° with the c axis. The inclination of α_{11} axis to the c axis as determined in the present study is 49°. The good agreement between the two indicates that the direction of maximum thermal expansion is very close to the normal to the

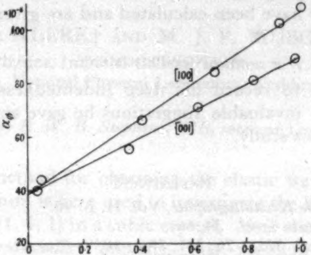


Figure 1. Variation of expansion coefficient (30–90°c) with $\sin^2 \phi$ in the two zones [100] and [001]. The circles represent the experimental values and the straight lines have been drawn using the calculated values of α_{a^*}, α_{c^*} and α_{22}.

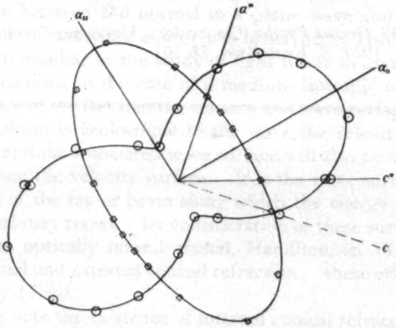


Figure 2. Variation of expansion coefficient (30–90°c) α_ξ and of $1/\sqrt{\alpha_\xi}$ with direction in the [010] zone. The dumb-bell shaped curve and the ellipse represent the theoretical variations to be expected. The circles and squares are the observed values.

oxygen planes. This is not surprising in view of the similar behaviour found in $NaNO_3$ and calcite, in both of which the direction of maximum expansion is perpendicular to the planes of O_3 groups.

From the values of the principal expansion coefficient it is possible to calculate the variation in the lattice dimensions of the crystal. The expansion coefficient along a, b and c axes (α_a, α_b, α_c) and the rate of change of the monoclinic angle with temperature $d\beta/dt$ have been calculated and are given in table 5.

ACKNOWLEDGMENT

The author wishes to record his deep indebtedness to Professor G. N. Ramachandran for the invaluable suggestions he gave and the keen interest he took in the course of this study.

REFERENCES

GROTH, P., 1910, *Chemische Kristallographie*, Vol. II, p. 90.
MADAN, H. G., 1886, *Nature, Lond.*, **34**, 66.
RAYLEIGH, LORD, 1888, *Phil. Mag.*, **26**, 241, 256; 1889, *Proc. Roy. Instn G.B.*, **12**, 447.
ROBERTSON, J. M., and UBBELOHDE, A. R., 1939, *Proc. Roy. Soc.* A, **170**, 241.
SHAW, R., 1953, *Acta Cryst., Camb.*, **6**, 428.
SIRKAR, S. C., 1930, *Indian J. Phys.*, **5**, 337.
SREEDHAR, A. K., 1952, *Proc. Indian Acad. Sci.* A, **36**, 141 ; 1953, *J. Indian Inst. Sci.*, **35**, 17.
UBBELOHDE, A. R., and WOODWARD, I., 1946, *Proc. Roy. Soc.* A, **185**, 448.
WHITTAKAR, E., and ROBINSON, G., 1944, *The Calculus of Observations* (London : Blackie), p. 243.
WOOSTER, W. A., 1949, *Crystal Physics* (Cambridge : University Press), p. 25.
ZACHARIASEN, W. H., 1929, *Z. Kristallogr.*, **71**, 501.

Internal Conical Refraction of Transverse Elastic Waves in a Cubic Crystal

By J. de KLERK[†] and M. J. P. MUSGRAVE[‡]

[†] Department of Physics, Imperial College, London S.W.7
[‡] Physics Division, National Physical Laboratory, Teddington, Middlesex

Communicated by R. W. B. Stephens; *MS. received* 12*th October* 1954

Abstract. A general method for obtaining the elastic wave surface appropriate to any aeolotropic medium is here used to investigate the behaviour of transverse plane waves of normal (1, 1, 1) in a cubic crystal. It is shown that the conditions for internal conical refraction prevail; this prediction has been verified experimentally.

§ 1. Introduction

The distinction between the normal to a plane wave and the direction of propagation of its associated energy was observed by Huygens in 1678 and has since been familiar in the study of light waves in crystals. The coincidence of these directions in the case of a medium isotropic to a given type of wave is a consequence of the fact that the velocity and wave surfaces are coincident spheres; if the medium is aeolotropic to the wave the velocity surface will no longer be spherical and the associated wave surface will also be non-spherical and different in form from the velocity surface. It is the wave surface which determines the direction of the ray or beam along which the energy associated with a wave of given normal may travel. By consideration of these surfaces appropriate to light waves in an optically biaxial crystal, Hamilton, in 1833, predicted the phenomena of internal and external conical refraction; these effects were experimentally verified by Lloyd.

In the following note the existence of internal conical refraction of transverse elastic waves of normal (1, 1, 1) is presaged, using a recent geometrical method for obtaining the elastic wave surface for any aeolotropic medium; an experimental investigation of this prediction by propagating elastic waves of ultrasonic frequency through a nickel crystal is described; the results verify the findings of the theory.

Analysis in progress at the time of writing indicates that the conditions for a cone of external refraction do not prevail; however, it is hoped to publish a detailed discussion of the behaviour of elastic waves in cubic crystals shortly. It may be mentioned here that any external refractive effects would be very difficult to demonstrate with the techniques at present available for the propagation and detection of ultrasonic transverse vibrations.

In this analysis: c_{11}, c_{12}, c_{44} are the elastic constants of a medium of cubic symmetry and $c_{11}{}^{s}$, $c_{12}{}^{s}$, $c_{44}{}^{s}$ the constants at saturation induction; ρ is the density of the medium and v the phase velocity of a plane wave.

§ 2. PROPAGATION OF WAVES IN A CUBIC CRYSTAL

The condition for non-zero displacement vectors of plane elastic waves of normal (l, m, n) in a medium of cubic symmetry may be written in the form

$$H^3 - aH^2 + c(a+b)\{m^2n^2 + n^2l^2 + l^2m^2\}H - c^2(a+2b)\,l^2m^2n^2 = 0 \quad \ldots\ldots(1)$$

where $H = \rho v^2 - c_{44}$, $a = c_{11} - c_{44}$, $b = c_{12} + c_{44}$, $c = c_{11} - c_{12} - 2c_{44}$.

This cubic yields three real roots $H_i (i = \mathrm{L}, \mathrm{T1}$ or $\mathrm{T2})$ corresponding with waves possessing displacement vectors of direction

$$(p_i : q_i : r_i) = \left[\frac{l}{H_i - l^2c} : \frac{m}{H_i - m^2c} : \frac{n}{H_i - n^2c}\right]. \quad \ldots\ldots(2)$$

The suffixes L, T1, T2 are used to denote the quasi-longitudinal and two quasi-transverse waves of normal (l, m, n).

It has been shown (Musgrave 1954) that the direction in which the energy of a given wave of normal (l, m, n) will be propagated in a general aeolotropic medium may be found by computing certain quantities and then using them to perform a simple construction.

For a medium of cubic symmetry these quantities are v, $c_{44}/\rho v$, (p, q, r), (λ, μ, ν) and $H/\rho v \cos\epsilon$ where

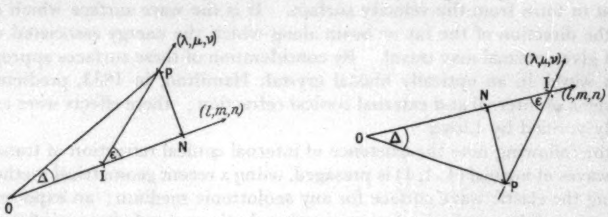
$$(\lambda, \mu, \nu) = \frac{\left(\dfrac{p^2}{l}, \dfrac{q^2}{m}, \dfrac{r^2}{n}\right)}{\left[\dfrac{p^4}{l^2} + \dfrac{q^4}{m^2} + \dfrac{r^4}{n^2}\right]^{1/2}} = \frac{(L, M, N)}{(L^2 + M^2 + N^2)^{1/2}} \quad \ldots\ldots(3)$$

and

$$\cos\epsilon = l\lambda + m\mu + n\nu. \quad \ldots\ldots(4)$$

In general, there are three sets of these quantities to be determined for each (l, m, n) corresponding with the three suffixes.

The construction is then performed as follows (see figure 1):



(a) $v > \dfrac{c_{44}}{\rho v}, \dfrac{H}{\rho v \cos\epsilon} > 0.$ Figure 1. (b) $v < \dfrac{c_{44}}{\rho v}, \dfrac{H}{\rho v \cos\epsilon} < 0.$

On the line of direction (l, m, n) through O mark off $\mathrm{ON} = v$ and $\mathrm{OI} = c_{44}/\rho v$; draw the line of direction (λ, μ, ν) making angle ϵ with (l, m, n) through I and mark off $\mathrm{IP} = H/\rho v \cos\epsilon$ paying due regard to the sign of this quantity. Then P is the point on the wave surface corresponding with N on the velocity surface and OP is the direction of the ray associated with a plane wave of velocity ON.

This correspondence is normally one to one; however the purpose here is to investigate and demonstrate a special case.

§ 3. PROPAGATION OF PLANE WAVES WITH NORMAL (1, 1, 1)

For this direction $l=m=n=1/\sqrt{3}$ and equation (1) may be factorized giving roots

$$H_L = \tfrac{1}{3}(a+2b) \quad\text{and}\quad H_{T1} = \tfrac{1}{3}c = H_{T2}.$$

Hence $p_L = q_L = r_L = 1/\sqrt{3} = \lambda_L = \mu_L = \nu_L$ and $\cos\epsilon_L = 1$, so that P coincides with N and the energy of the wave is propagated along the wave normal.

For transverse waves we find that (p_T, q_T, r_T), are indeterminate, but by the proof for the orthogonality of displacement vectors of different velocities (Kelvin 1904), they must lie in the plane of the wave. Now from any two such directions which form an orthogonal triad with (1, 1, 1) we may form an arbitrary displacement vector contained by the plane of normal (1, 1, 1). Thus choosing the set

$$\frac{1}{\sqrt{3}} : \frac{1}{\sqrt{3}} : \frac{1}{\sqrt{3}} \quad \text{L} \qquad \frac{1}{\sqrt{2}} : -\frac{1}{\sqrt{2}} : 0 \quad \text{T1} \qquad \frac{1}{\sqrt{6}} : \frac{1}{\sqrt{6}} : -\frac{2}{\sqrt{6}} \quad \text{T2}$$

we may define any transverse wave vector by

$$p_T = \frac{1}{\sqrt{2}}\cos\psi + \frac{1}{\sqrt{6}}\sin\psi, \qquad \ldots\ldots(5)$$

$$q_T = -\frac{1}{\sqrt{2}}\cos\psi + \frac{1}{\sqrt{6}}\sin\psi, \qquad \ldots\ldots(6)$$

$$r_T = -\frac{2}{\sqrt{6}}\sin\psi, \qquad \ldots\ldots(7)$$

where ψ is an arbitrary angle in the plane of the wave measured from $(1, \bar{1}, 0)$, whence

$$L_T = \frac{\sqrt{3}}{2}\left[\cos^2\psi + \frac{\sin^2\psi}{3} + \frac{2}{\sqrt{3}}\cos\psi\sin\psi\right] = \frac{1}{\sqrt{3}}\left[1+\cos\left(2\psi - \frac{\pi}{3}\right)\right], \quad \ldots\ldots(8)$$

$$M_T = \frac{\sqrt{3}}{2}\left[\cos^2\psi + \sin^2\psi - \frac{2}{\sqrt{3}}\cos\psi\sin\psi\right] = \frac{1}{\sqrt{3}}\left[1+\cos\left(2\psi + \frac{\pi}{3}\right)\right], \quad \ldots\ldots(9)$$

$$N_T = \frac{\sqrt{3}}{2}\left[\frac{4\sin^2\psi}{3}\right] = \frac{1}{\sqrt{3}}\left[1-\cos 2\psi\right] \qquad \ldots\ldots(10)$$

and we may notice that for all values of ψ, (L_T, M_T, N_T), which defines the same direction as $(\lambda_T, \mu_T, \nu_T)$, is perturbed from the value $1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}$.

Using these values in equations (3) and (4) it follows that $\cos\epsilon_T = \sqrt{(2/3)}$ for all ψ.

Thus as the transverse displacement vector varies with ψ, the direction $(\lambda_T, \mu_T, \nu_T)$ rotates about (1, 1, 1) generating a cone of semi-angle $\epsilon_T = \cos^{-1}\sqrt{(2/3)} = 35{\cdot}4°$. From the form of (8), (9) and (10) it will be seen that one such complete rotation occurs as ψ varies over a range of π radians.

Referring to figure 1 it may be seen that as (p_T, q_T, r_T) varies with ψ, the point P describes a circle and hence there exists a cone of possible rays associated with different displacement vectors. It is therefore possible to predict the phenomenon of *internal conical refraction* for plane elastic transverse waves of normal (1, 1, 1) in a medium of cubic symmetry. The semi-angle of this cone of refraction Δ_c is given by

$$\tan\Delta_c = \left|\frac{H_T}{\rho v_T^2}\right| \tan\epsilon_T. \qquad \ldots\ldots(11)$$

On substitution of the appropriate values of H_T and v_T in terms of the elastic constants this becomes

$$\tan \Delta_c = \left| \left[1 - \frac{3c_{44}}{c_{11} - c_{12} + c_{44}} \right] \right| \frac{1}{\sqrt{2}} = \left| \frac{c}{\sqrt{2}(c + 3c_{44})} \right|. \quad \ldots\ldots (12)$$

It may be noted that this angle is dependent upon $c = c_{11} - c_{12} - 2c_{44}$ which is the normal measure of the degree of elastic aeolotropy possessed by a cubic crystal. When $c = 0$, as for isotropic materials, then $\Delta_c = 0$ and there is no cone of refraction.

The elastic constants of unmagnetized nickel as determined by one of the authors (J. de K.) are $c_{11} = 24{\cdot}3_6$, $c_{12} = 14{\cdot}9_4$, $c_{44} = 11{\cdot}9_6$ (all $\times 10^{11}$ dyn cm^{-2}).

These values give $\Delta_c = 25°$ for nickel; the values for several other metals are: W, 0; Al, 5; Ni, 25; Fe, 25; Cu, 27.

§ 4. PRACTICAL CONSIDERATIONS OF THE PROBLEM

To show experimentally the existence of internal conical refraction requires as large a single crystal as possible and correspondingly the narrowest possible ultrasonic beam. The authors were successful in obtaining a single crystal of nickel with the following axis dimensions: $l_{[001]} = 1{\cdot}92$ cm, $l_{[110]} = 1{\cdot}87$ cm, $l_{[111]} = 1{\cdot}80$ cm and to achieve the second condition 4 mm × 6 mm Y cut quartz crystals were employed as transmitter and detector respectively.

Since the beam spread is proportional to the ratio of the elastic wavelength λ in the crystal to the diameter of the transducer, it was desirable to employ a high ultrasonic frequency, i.e. small wavelength. The upper limit of this frequency is set by the increasing attenuation of the elastic waves in the medium as the frequency is raised, and so a compromise had to be adopted, 5 Mc/s being taken as the optimum frequency. The rectangular form of the transducers permitted the easier identification of their directions of polarization, which were along one diagonal.

The most desirable coupling medium between the transducers and the nickel crystal was one which allowed good transmission for shear waves while permitting easy movement of the transducers over the nickel surface. Salol (phenyl salicylate) was chosen for its low melting point (45°c) and its excellent shear wave transmission characteristic.

§ 5. PREPARATION OF NICKEL SINGLE CRYSTAL

The single crystal, cut from an ingot, had surfaces parallel to ⟨001⟩, ⟨110⟩ and ⟨111⟩ planes. The surfaces were prepared by grinding on a high speed emery disc, using alcohol as a lubricant, and the disturbed surface layer was removed by etching in a solution containing ammonium persulphate, potassium cyanide, ammonia and distilled water. The orientations of the final faces, determined by electron diffraction by Dr. H. Wilman, were found to be within 1° of the desired planes.

§ 6. TRANSDUCER SYSTEM

Both surfaces of the transducers were gold plated and the backings were brass plates 3 mm thick, on to which were soldered short lengths of 40 gauge enamelled copper wire, to form conductors to the transmitter and to the receiver respectively. The nickel specimen acted as the earth return for both transducers. In practice the position of the transmitter transducer was left undisturbed, while that of the receiver was moved approximately 1 mm after each measurement of received pulse amplitude. This necessitated melting the salol by heating the nickel crystal before the position of the receiver transducer could be altered, and then recrystallizing the salol by cooling.

§ 7. TRANSDUCER CIRCUIT

A crystal controller synchronizer provided 1 μsec pulses at 1 μsec and 6 μsec intervals and 20 μsec pulses at a repetition rate of 1200 or 600 per second. The 6 μsec and 20 μsec pulses were fed (figure 2) to the hard valve transmitter which provided 1 μsec radio-frequency pulses whose frequency could be varied from 2 Mc/s to 30 Mc/s. Ultrasonic elastic waves were introduced into the nickel crystal by applying these radio-frequency pulses to the transmitter transducer, which was cemented to the crystal.

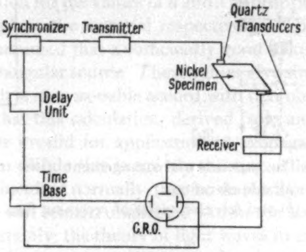


Figure 2. Block diagram of the apparatus.

The 1 μsec, 6 μsec and 20 μsec pulses from the synchronizer were fed to an electronic delay network, the output of which was used to trigger a miller integrator time base generator. The delay network delayed the start of the time base, in fixed steps of 1 μsec or 6 μsec, after the start of the transmitter pulse. By this means the velocity of the elastic waves in the nickel crystal was measured.

The arrival of the elastic waves at the opposite surface of the nickel crystal excited the receiver transducer and its electrical output was amplified by a wide-band radio-frequency amplifier and displayed on a cathode-ray tube.

§ 8. MEASUREMENT OF ANGLE OF REFRACTION Δ_c

The positions of the transmitter and receiver transducers on the two $\langle 111 \rangle$ faces were plotted by means of a travelling microscope, and the angle Δ calculated from these measurements using the relation $\tan \Delta = d/l$ (see figure 3). For these calculations, the centre of each transducer was regarded as defining its position.

The transmitter transducer was first placed in the centre of one of the $\langle 111 \rangle$ faces, with its direction of polarization parallel to the $\langle 110 \rangle$ faces. The receiver transducer was placed at the centre of the opposite $\langle 111 \rangle$ face and rotated until the received signal amplitude was a maximum, to obtain its correct direction of polarization. Retaining this polarization the receiver transducer was moved along the surface in the $[\bar{1}\bar{1}2]$ direction (see figure 3 (*b*)) while both the amplitude and velocity of the detected elastic wave were measured at 1 mm intervals. The receiver transducer was then placed on this line at the position of maximum received signal. The distance between transmitter and receiver transducers in their positions for maximum received signal was d_c. The receiver transducer was next moved across the $\langle 1, 1, 1 \rangle$ surface along an arc of a circle whose centre was that of the transmitter and whose radius was d_c; the amplitude and velocity of the received elastic wave were measured as before. Plots of the readings taken are shown in figure 4 (*a*) and (*b*); they indicate that $\Delta_c = 21°$ (cf. 25° from theory).

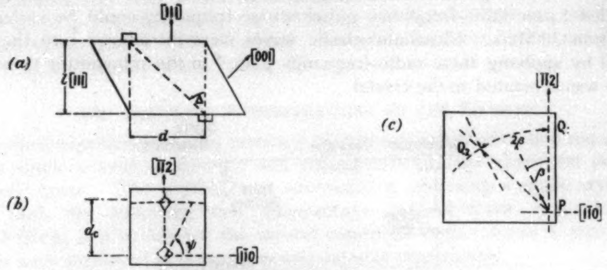


Figure 3. The experimental arrangement of the transducers :
(*a*) side elevation,
(*b*) and (*c*) plan views of receiving face.

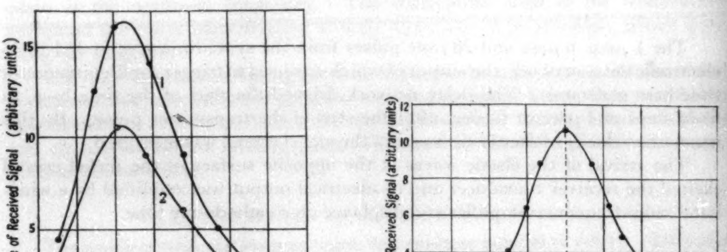


Figure 4. Variation of signal amplitude with (*a*) angle of deviation from wave normal, Δ, measured in [11̄0] plane (see figure 3 (*a*)), and (*b*) angle ψ (see figure 3 (*b*) and analysis).

The transit time of energy was found to be the same for all path lengths corresponding with different values of Δ. This indicated that the waves possessed a common phase velocity and that no energy arriving at the receiver transducer had been internally reflected.

§ 9. ANGLE OF DIVERGENCE OF THE BEAM

The width of the response curves in figure 4 is due to the spread of the beam of elastic waves. Whenever the linear dimensions of a source of vibrational energy comprise a large number of wavelengths the energy of the wave motion has the directional character of a group of rays forming a narrowly divergent beam.

A calculation of the displacement field within an isotropic solid arising from a circular source of radius a, *small* compared with the wavelength λ, has recently been given by Miller and Pursey (1954). From their analysis it appears that for sources *larger* than a wavelength each amplitude $A(\theta)$ of the displacement field obtained from a small source should be multiplied by a factor $2J_1(x)/x$ where $x = 2\pi a \sin\theta/\lambda$.

To obtain an approximate value of the beam angle in the experiments presently described, it is thus sufficient to determine the value of $\sin\theta$ which gives the first zero of the Bessel function for the values of a and λ appropriate to the size of source and acoustic wavelength in the material respectively. For the waves examined $\lambda = 0{\cdot}056$ cm and it is assumed that a sufficiently good value of a is that for a disc of area equal to the rectangular source. These values give $\sin\theta = 3{\cdot}83\lambda/2\pi a = 0{\cdot}122$, and $\theta = 7^\circ$. This result is in reasonable accord with that observed (cf. figure 4).

It may be argued that this calculation, derived from an analysis of conditions in an isotropic solid, is invalid for application to propagation in an aeolotropic medium like a crystal. While this is strictly the case, it is reasonable to assume that though the beam, incident normally, may be markedly refracted by the crystal, its angle of divergence will remain unaffected within the accuracy of this approximate calculation. Certainly, the theory of light waves in crystals, with which the theory of elastic waves has strong qualitative analogy, suggests the validity of this view.

§ 10. EFFECT OF MAGNETIC FIELD ON Δ_c

When the nickel crystal is placed in a magnetic field the elastic constants c_{11}, c_{12}, c_{44} change, and hence both the velocity and Δ_c change. The values of the elastic constants at saturation induction, c_{11}^s, c_{12}^s, c_{44}^s, were calculated from the longitudinal and two shear velocities in the $[1\bar{1}0]$ direction and found to be $c_{11}^s = 24{\cdot}6_5$, $c_{12}^s = 14{\cdot}7_3$, $c_{44}^s = 12{\cdot}4_7$ (all $\times 10^{11}$ dyn cm^{-2}). These show $\Delta_c^s - \Delta_c \sim 0{\cdot}5^\circ$.

At 5 Mc/s this change in Δ_c is too small for measurement by the present technique. (This estimate has assumed that cubic symmetry is preserved on magnetization.)

§ 11. INTERNAL CONICAL REFRACTION

The size and geometry of the nickel crystal did not allow a complete circular section of the cone of refraction to be observed but a conclusive demonstration of the phenomena associated with the effect was obtained in the following way.

From positions of maximum received signal P and Q_1 (figure 3 (*c*)) the common polarization of transmitter and receiver was rotated through an angle β; the receiver was then moved to regain maximum signal. Traverse along the arc Q_1Q_2 (figure 4 (*b*)) and the radius PQ_2 (figure 4 (*a*)) defined P and Q_2 as the required positions; Q_1PQ_2 was found to be 2β, viz. twice the angle of rotation of the displacement vector as was indicated by equations (8), (9) and (10). This same result was observed for several values of the angle β. It is concluded that on a sufficiently large specimen a complete circular section of the cone of refraction would be traced as the displacement vector was rotated through 180°.

ACKNOWLEDGMENTS

The analysis which is here presented, and which suggested the foregoing experimental observations, is a small section of a more general investigation which forms part of the research programme of the National Physical Laboratory. The experimental work was performed in the Physics Department of Imperial College. Both authors wish to acknowledge helpful discussions with Dr. R. W. B. Stephens. The paper is published with the approval of the Director of the National Physical Laboratory.

REFERENCES

HAMILTON, SIR W. R., 1931, *Collected Papers*, Vol. I (Cambridge : University Press).
KELVIN (LORD), 1904, *Baltimore Lectures* (London : Oxford University Press).
LLOYD, REV. H., 1833, *Proc. R. Irish Acad.*, p. 145.
MILLER, G. F., and PURSEY, H., 1954, *Proc. Roy. Soc.* A, **223**, 521.
MUSGRAVE, M. J. P., 1954, *Proc. Roy. Soc.* A, **226**, 339.

Some Applications of Fourier Series in the Numerical Treatment of Linear Behaviour

By F. C. ROESLER

Imperial Chemical Industries Limited, Butterwick Research Laboratories, Welwyn, Herts.

MS. received 12th July 1954

Abstract. Earlier work is here continued. Certain functional relations common to various phenomenological theories, like the theory of linear visco-elasticity and the theory of linear electrical networks, may be written as integral equations of the convolution type. As the circular functions are eigenfunctions of the convolution type kernels, such equations can be solved by Fourier expansion. A number of possible applications of such Fourier expansions are indicated and the relevant details are worked out. The method centres on the evaluation of the eigenvalues for each kernel. The examples discussed refer to the theory of visco-elasticity and include the determination of the relaxation spectrum either from the dynamic elasticity as a function of frequency or from the stress relaxation as a function of time.

§ 1. INTRODUCTION

This paper continues and extends the work reported in an earlier publication (Roesler and Pearson 1954).

Fourier series are commonly used in the description of linear, time dependent phenomena as a means of representing variables such as stresses, voltages, velocities, etc. The basis of this representation lies in the existence of proper modes, where the variables change sinusoidally in time, and sometimes also in space. The use of Fourier series described here is rather different. These series will represent functions which describe intrinsic properties of the system under consideration, for instance the dependence of impedance on frequency. The independent variables in these functions will be the logarithms of certain ratios of times or frequencies.

For definiteness, the terminology of the theory of linear visco-elasticity will be used in what follows. The monograph by Gross (1953) gives a very full exposition of the mathematical structure of this theory. The integral equations considered below occur there and also in other sources. Formally identical equations occur in the theories of other linear phenomena.

All the properties which make up the linear visco-elastic behaviour of a given substance may be defined by specifying a single function. One such function is the relaxation spectrum, which is a certain distribution or density function. The spectral density is a quantity which cannot be measured directly with any known practical apparatus. Its indirect evaluation from empirical data involves the solution of an integral equation of the first kind.

The equations which link the relaxation spectrum to those properties of the substance which are of most general interest, and which are most easily accessible experimentally, can all be written in a form known as the convolution type. Under certain conditions, such equations may be solved by means of Fourier integrals (Titchmarsh 1937). One may regard this method as a solution by expansion into eigenfunctions. The eigenvalue spectrum of the relevant kernels is continuous, but for certain purposes (for numerical work) it may be replaced by a discrete one, and then the Fourier integrals become Fourier series. The main step in effecting the solution of any convolution type integral equation of the first kind,

$$f(s)=\int_{-\infty}^{+\infty} K(s-t)\,g(t)\,dt, \qquad \ldots\ldots(1)$$

is to find the eigenvalues of the kernel $K(s-t)$. It can be shown that the eigenvalue $\lambda(\nu)$ associated with the eigenfunction $\exp(-i\nu s)$ is given by

$$\frac{1}{\lambda(\nu)}=\int_{-\infty}^{+\infty} K(\xi)\exp(i\nu\xi)\,d\xi, \qquad \ldots\ldots(2)$$

provided the integral exists.

In the preceding paper, the solution by means of Fourier expansion of one special integral equation was described. Some other equations will now be treated. A number of general questions, as the existence of the solution, the replacing of the Fourier integrals by series, the truncation of such series, the elimination of noise and the resolution power, will not be treated again here. The discussion would follow similar though not identical lines for every new equation. No numerical examples are given in this paper (the one given previously is fairly representative), but the methods described have all been tried numerically and have been found workable.

§ 2. SPECTRUM FROM DYNAMIC ELASTICITY

The real part of the complex modulus, also called the dynamic elasticity, is linked to the relaxation spectrum by the equation

$$E'(\Omega)=\int_0^{\infty}\phi(T)\frac{\Omega^2T^2}{1+\Omega^2T^2}\,dT. \qquad \ldots\ldots(3)$$

Here Ω stands for the angular frequency and $\phi(T)$ is the relaxation spectrum referred to a linear time scale. On introducing logarithmic variables $\tau=\ln(T/T_0)$ and $\omega=-\ln(\Omega T_0)$, and with a re-definition of the spectrum, such that it is now a density function with respect to τ, equation (3) becomes

$$E'(\omega)=\tfrac{1}{2}\int_{-\infty}^{+\infty}[1+\tanh(\tau-\omega)]\,\phi(\tau)\,d\tau. \qquad \ldots\ldots(4)$$

This is a convolution-type equation. However, as the kernal is not integrable, its Fourier transform is not well defined, and the relation (2) for finding the eigenvalues cannot be applied straightforwardly. One may overcome this difficulty by considering the differentiated form of (4), viz.:

$$-\frac{dE'(\omega)}{d\omega}=\frac{1}{2}\int_{-\infty}^{+\infty}\operatorname{sech}^2(\tau-\omega)\,\phi(\tau)\,d\tau. \qquad \ldots\ldots(5)$$

The relation (2) can now be applied. One finds

$$\frac{1}{2}\int_{-\infty}^{+\infty}\operatorname{sech}^2\exp(i\nu\xi)\,d\xi=\frac{\nu\pi}{2\sinh(\nu\pi/2)} \qquad \ldots\ldots(6)$$

(cf. Campbell and Foster (1942) pair 614).

Thus the eigenvalues of (5) are given by

$$\lambda(\nu) = \frac{2}{\nu\pi}\sinh(\nu\pi/2). \qquad \ldots\ldots(7)$$

The Fourier expansion of the left side in (5) may be found from the Fourier expansion of $E'(\omega)$ itself, term by term, thus eliminating the necessity of a numerical differentiation. $E'(\tau)$ is a monotonically decreasing function and it can be represented by a cosine series within the relevant range. Thus one may put

$$E'(x) = \sum_0^n c_k \cos kx, \qquad \ldots\ldots(8)$$

where x is a variable linearly related to τ, and chosen such that the range of τ which is of physical interest falls within $0 < x < \pi$. For the left side in (5) one finds

$$-\frac{dE'}{d\omega} = -\frac{dE'}{dx}\frac{dx}{d\omega} = \frac{\pi}{L}\sum_1^n kc_k \sin kx. \qquad \ldots\ldots(9)$$

Here L is the length of the domain of τ which corresponds to $0 < x < \pi$. The spectrum is now given by a sine series in x:

$$\phi(x) = \sum_1^n b_k \sin kx. \qquad \ldots\ldots(10)$$

The b_k are given by

$$b_k = \lambda_k \frac{\pi}{L} kc_k. \qquad \ldots\ldots(11)$$

Using (7) and taking account of the change from τ to x, one obtains

$$b_k = \frac{2c_k}{\pi}\sinh\left(\frac{k\pi^2}{2L}\right). \qquad \ldots\ldots(12)$$

The problem of evaluating $\phi(\tau)$ from $E'(\omega)$ is thus solved.

§ 3. ADJUSTMENT OF DIFFERENT EMPIRICAL FUNCTIONS

Supposing that both $E'(\omega)$ and $E''(\omega)$ have been determined experimentally for the same substance, the relaxation spectrum can then be evaluated from either set of data. If both computations were carried out the two results should agree. In practice, there will be some disagreement. By adopting a single, intermediate spectrum, one may hope, working back from this, to obtain more reliable empirical functions. This adjustment can in fact be carried out without explicitly finding the spectrum, and the direct process is not only simpler but also more accurate.

Both E'' and E' may be made functions of the same variable x, a linear function of τ chosen as explained above. Both E'' and E' may then be expressed by Fourier series in x, and these will read

$$E''(x) = \sum_1^n a_k \sin kx, \quad \ldots\ldots(13); \qquad E'(x) = \sum_0^n c_k \cos kx. \quad \ldots\ldots(14)$$

The spectrum is given by

$$\phi(x) = \sum_1^n b_k \sin kx. \qquad \ldots\ldots(15)$$

The b_k can be related to either a_k or c_k. Thus one has

$$b_k = \frac{2a_k}{\pi}\cosh\left(\frac{k\pi^2}{2L}\right) = \frac{2c_k}{\pi}\sinh\left(\frac{k\pi^2}{2L}\right). \qquad \ldots\ldots(16)$$

Therefore the direct relation between a_k and c_k is

$$a_k = c_k \tanh(k\pi^2/2L). \qquad \ldots\ldots(17)$$

As the hyperbolic tangent does not increase beyond 1, the comparison of a_k and c_k afforded by relation (17) is not disturbed by the amplification of the noise which tends to occur in the determination of the spectrum.

Equation (17) may be used to find one of the functions $E''(\omega)$ and $E'(\omega)$ if only the other has been determined experimentally. Relations similar to (17), which are useful either for the adjustment of two functions or for the evaluation of one function from the other, may be set up for any other pairs of functions generated from $\phi(\tau)$ by means of convolution-type integrals.

§ 4. CASE OF GIVEN STRESS RELAXATION FUNCTION

A specimen of material exhibiting linear stress relaxation is deformed at time $t=0$ by suddenly imposing some prescribed, uniform and thereafter constant strain. Oscillations, which might be caused by inertia effects, are ignored. The stress in the material will then at later times be given by

$$\sigma(t) = \int_0^\infty \phi(T)\exp(-t/T)\,dT. \qquad \ldots\ldots(18)$$

The function $\phi(T)$ is the spectrum of relaxation times referred to a linear time scale, and is identical with the $\phi(T)$ in (3).

If a specimen of material exhibiting linear, recoverable creep has been loaded by imposing a prescribed, uniform, constant stress for an infinite time, the strain which has thereby been caused will, after sudden unloading at time $t=0$, decrease according to

$$\epsilon(t) = \int_0^\infty \chi(T)\exp(-t/T)\,dT. \qquad \ldots\ldots(19)$$

Again, the oscillations which might follow the sudden releasing of the stress are left out of account. The function $\chi(T)$ in (19) is called the retardation spectrum of the material.

It is seen that (18) and (19) are formally the same integral equation. On putting $\tau = \ln(t/T_0)$ and $\omega = \ln(T/T_0)$, and with the re-definition of the spectrum as before, equation (18) becomes

$$\sigma(\tau) = \int_{-\infty}^{+\infty} \phi(\omega)\exp[-\exp(\tau-\omega)]\,d\omega. \qquad \ldots\ldots(20)$$

This is of the desired convolution type, but as in § 2 it is preferable to consider the differentiated form of (20), since the kernel of (20) itself is not integrable. Differentiation of (20) gives

$$-\frac{d\sigma(\tau)}{d\tau} = \int_{-\infty}^{+\infty} \phi(\omega)\exp[\tau-\omega-\exp(\tau-\omega)]\,d\omega. \qquad \ldots\ldots(21)$$

The kernel of (21) is given by

$$K(\xi) = \exp[\xi - \exp\xi]. \qquad \ldots\ldots(22)$$

This is a smooth and integrable function, and therefore its Fourier transform exists. However, it is not readily expressible in terms of known functions. Therefore the consequences of the rule (2) for the eigenvalues must, in the present case, be worked out numerically. As this can be done once and for all, no real difficulty arises.

It is best to choose, at this stage, one given basic domain for all the Fourier analysis and synthesis procedures connected with the solving of equations (20) and (21). This domain is characterized by a certain length L, in terms of the variables τ, ω and $\tau-\omega$, which corresponds to the length π in terms of the variable x. All Fourier integrals, including that occurring in (2), are now replaced by Fourier series. The rule for the eigenvalues, formerly (2), now becomes

$$\frac{1}{\lambda_k}=\frac{L}{\pi}\int_{-\infty}^{+\infty} K(x)\exp(ikx)\,dx. \qquad \ldots\ldots(23)$$

This gives the eigenvalue associated with the eigenfunction $\exp(-ikx)$. The kernel itself may be represented by a Fourier series in x

$$K(x)=\sum_{-\infty}^{+\infty}\gamma_k\exp(-ikx). \qquad \ldots\ldots(24)$$

The relation of the complex coefficients γ_k to the ordinary cosine and sine coefficients will be given later. Inserting (24) into (23) and taking account of the orthogonality and the normalization properties of the circular functions, there results

$$1/\lambda_k=2L\gamma_k. \qquad \ldots\ldots(25)$$

If then the left side of (21) is developed into a series

$$-\frac{d\sigma}{d\tau}=\sum_{-n}^{+n}C_k\exp(-ikx), \qquad \ldots\ldots(26)$$

the coefficients of a corresponding series which gives the spectrum are

$$S_k=\lambda_kC_k=C_k/(2L\gamma_k). \qquad \ldots\ldots(27)$$

This result must now be expressed in terms of the ordinary cosine and sine coefficients of all the functions. The notation is as follows:

Function	Complex coefficient	Cosine coefficient	Sine coefficient
$K(x)$	γ_k	α_k	β_k
$-d\sigma/d\tau$	C_k	A_k	B_k
$\phi(x)$	S_k	P_k	R_k

The relations between γ_k, α_k, β_k are given by

$$\gamma_k=\begin{cases}\frac{1}{2}(\alpha_k+i\beta_k), & k>0\\ \frac{1}{2}(\alpha_{|k|}-i\beta_{|k|}), & k<0\end{cases} \qquad \ldots\ldots(28)$$

$$\alpha_k=\gamma_k+\gamma_{-k}, \qquad i\beta_k=\gamma_k-\gamma_{-k}. \qquad \ldots\ldots(29)$$

The relations of the other two triples are analogous. Using these, one finds from (27)

$$P_k=\frac{1}{L}\,\frac{A_k\alpha_k+B_k\beta_k}{\alpha_k^2+\beta_k^2} \quad \ldots\ldots(30); \qquad R_k=\frac{1}{L}\,\frac{B_k\alpha_k-A_k\beta_k}{\alpha_k^2+\beta_k^2}. \quad \ldots\ldots(31)$$

Finally, the differentiation which leads from σ to $-d\sigma/d\tau$ must be considered. As in § 2, the differentiation can be done on a Fourier series. A difficulty is met with here. The domain of x used in the present section for the representation of the given set of data runs from $x=-\pi$ to $x=+\pi$ (not, as in other sections, from $x=0$ to $x=\pi$). The use of a complete period is necessary because the kernel of (21) is non-symmetric. This, however, entails that $\sigma(x)$ cannot now be regarded as a periodic function without it having a jump at $x=-\pi$ and $x=+\pi$ (within this domain $\sigma(x)$ decreases monotonically). Such a jump is not permissible in any function the derivative of which shall be a 'possible' left-hand side in (21). In practice, implicit introduction of this jump would lead to poor convergence of the series. The jump must therefore be avoided. This may be done by subtracting from the given $\sigma(\tau)$ a linear function $\eta(\tau)$ defined such that the difference

$$\sigma(\tau) - \eta(\tau) = \sigma^*(\tau) \qquad \text{......(32)}$$

vanishes for those values of τ which correspond to $x = -\pi$ and $x = \pi$. This difference σ^* is now regarded as a function of x, and $\sigma^*(x)$ may be assumed to possess a periodic continuation, without thereby bringing in jumps. On the other hand, as $\eta(\tau)$ is linear, the derivative of σ^* differs from that of σ by a constant only. This constant is, from the choice of η, given by

$$-\frac{d\eta}{d\tau} = \frac{\sigma_{-\pi} - \sigma_{+\pi}}{2L}. \qquad \text{......(33)}$$

Here $\sigma_{-\pi}$ and $\sigma_{+\pi}$ denote the values of $\sigma(x)$ for $x = -\pi$ and for $x = +\pi$. The eigenvalue of (21) associated with a constant is 1, therefore (33) goes over into the spectrum as a constant term, without a change of amplitude.

The function $\sigma^*(x)$ is analysed in terms of a mixed (cosine and sine) Fourier series:

$$\sigma^*(x) = \sum_0^n A_k^* \cos kx + \sum_1^n B_k^* \sin kx. \qquad \text{......(34)}$$

From this

$$-\frac{d\sigma^*}{d\tau} = -\frac{d\sigma^*}{dx}\frac{dx}{d\tau} = \sum_1^n \frac{\pi}{L} k A_k^* \sin kx - \sum_1^n \frac{\pi}{L} k B_k^* \cos kx. \qquad \text{......(35)}$$

This is to be compared with the development given by (26) and the expression of the C_k in terms of A_k, B_k. One identifies

$$A_k = -\frac{\pi}{L} k B_k^* \quad \text{.....(36);} \qquad B_k = \frac{\pi}{L} k A_k^*. \quad \text{......(37)}$$

Rewriting equations (30) and (31) in terms of the starred variables one obtains

$$P_k = \frac{k\pi}{L^2} \frac{A_k^* \beta_k - B_k^* \alpha_k}{\alpha_k^2 + \beta_k^2} \quad \text{......(38);} \qquad R_k = \frac{k\pi}{L^2} \frac{A_k^* \alpha_k + B_k^* \beta_k}{\alpha_k^2 + \beta_k^2}. \quad \text{......(39)}$$

The spectrum is then given by

$$\phi^*(x) = \sum_1^n P_k \cos kx + \sum_1^n R_k \sin kx. \qquad \text{......(40)}$$

The star in $\phi^*(x)$ is a reminder of the change from $\sigma(x)$ to $\sigma^*(x)$. To get the correct spectral density ϕ, the constant given by (33) must be added to ϕ^*.

In order to use the formulae (38) and (39) for the evaluation of a relaxation spectrum, one requires the α_k, β_k coefficients of the kernel (22). Their values depend upon the length L, and L must be chosen with a view to the range and accuracy of the data to be analysed. (The more accurate the data, the larger L must be chosen, to make the implicit assumption of periodicity harmless.) In the table a set of values is given for α_k, β_k and $k\pi/[L^2(\alpha_k^2 + \beta_k^2)]$. These were computed for an L of $12{\cdot}8\tau$ units length and if they are to be used the Fourier coefficients A_k^* and B_k^* must be evaluated for a domain with the same L. The total range $-\pi < x < +\pi$ corresponds to $2L$ and with $L = 12{\cdot}8$ units of τ it is $2 \times 12{\cdot}8/\ln 10 \simeq 11$ decades long. This will be more than adequate in most cases. If the data cover a less extended range only (but do tend to a limiting, constant value at either end of the observed range) one may still use the larger domain in the analysis.

§ 5. CONCLUDING REMARKS

The work reported in the preceding and in the present paper was started as an attempt to further the theory of visco-elastic behaviour, and therefore it is relevant to consider its potential usefulness. This will concern only the special field mentioned, and not the possibility of application in other fields.

In the theory of visco-elasticity, the relaxation spectrum (or the retardation spectrum) is of interest for two reasons. Firstly, it provides a convenient description of the sum total of mechanical properties. Secondly, there is a general expectation that the spectrum (rather than the more easily determined properties) can be correlated with the molecular structure of the material. Thus knowledge of the spectrum should ultimately further the real understanding of the properties of substances like high polymers.

Fourier coefficients α_k and β_k of the kernel $K(x)$ defined by $K(\xi)$ of relation (22) within $-12{\cdot}8<\xi<12{\cdot}8$ and by periodic continuation outside of this domain, which is identified with $-\pi<x<\pi$. Four figure values of 129 ordinates of K were used in the evaluation of the tabled coefficients

k	$10^3\alpha_k$	$10^3\beta_k$	$\dfrac{\pi k}{L^2(\alpha_k^2+\beta_k^2)}$	k	$10^3\alpha_k$	$10^3\beta_k$	$\dfrac{\pi k}{L^2(\alpha_k^2+\beta_k^2)}$
0	78·14	0	0	9	8·68	2·575	2105
1	73·75	−10·084	3·4607	10	5·69	3·120	4553
2	63·11	−15·511	9·0801	11	3·43	3·11	9840
3	50·77	−15·591	20·394	12	1·80	2·76	21200
4	39·67	−12·408	44·395	13	0·69	2·24	45400
5	30·52	− 8·123	96·119	14	0	1·64	99300
6	23·17	− 4·095	207·81	15	−0·35	1·10	218000
7	17·26	− 0·903	449·3	16	−0·48	0·65	467000
8	12·49	1·300	973				

Regarding this second and more ambitious role that the spectrum might play, the immediate prospects of the methods developed in the two papers do not seem to the author to be very promising. Unfortunately, the interpretation of the spectra is so difficult, and in so rudimentary a stage, that at the present moment one does not really gain much by knowing the shape of a certain peak in the spectrum more accurately than as given by the damping function itself. It might be suggested, of course, that more exact knowledge of the spectra will make their interpretation easier. Although this may well ultimately come about, it would not seem that much optimism in this matter is warranted.

A more immediately attractive application is to problems such as the adjustment of empirical functions and the evaluation of one property of a substance from another. As discussed in § 3, the spectrum need not be found explicitly for such purposes. This kind of problem can be handled very well by means of Fourier series, and the practical accuracy is high.

To return to the determination of the spectrum, it may be mentioned that in the case of some of the relevant equations this problem admits of a different approach, which is sometimes likely to be more satisfactory. It is an iteration procedure which is based on the Fourier method in principle and yet in which no Fourier analysis need explicitly be made. This alternative method will be described in a separate paper by W. A. Twyman and the present author. It is not, unfortunately, applicable to the equation treated here in § 4, which is the most troublesome one.

ACKNOWLEDGMENTS

The author wishes to thank Dr. H. Kolsky and Mr. J. J. Benbow for many profitable discussions, and Mr. W. A. Twyman for assisting him in trial computations and in the work for the table.

REFERENCES

CAMPBELL, G. A., and FOSTER, R. M., 1942, *Fourier Integrals for Practical Applications* (New York : Van Nostrand).

GROSS, B., 1953, *Mathematical Structure of the Theories of Viscoelasticity* (Paris : Hermann).

ROESLER, F. C., and PEARSON, J. R. A., 1954, *Proc. Phys. Soc.* B, **67**, 338.

TITCHMARSH, E. C., 1937, *Introduction to the Theory of Fourier Integrals* (Oxford : Clarendon Press).

An Iteration Method for the Determination of Relaxation Spectra

By F. C. ROESLER and W. A. TWYMAN
Imperial Chemical Industries Ltd., Butterwick Research Laboratories, Welwyn, Herts.

MS. received 12*th July* 1954 *and in final form* 6*th December* 1954

Abstract. The determination of relaxation spectra involves the numerical solution of certain linear, convolution-type integral equations of the first kind. A simple iteration method is shown by which some such equations may be solved. In the limit this is equivalent to the exact solution by means of Fourier integrals. If only moderate accuracy is required, the iteration method is sometimes less laborious than the use of Fourier series, proposed earlier, and in any case it involves only elementary arithmetical operations.

As one illustration, the relaxation spectrum of polyisobutylene is found from the damping, for the whole domain over which published data are available. In the rubber–glass transition region the results agree well with those previously obtained for this region by use of Fourier series. As a second illustration, successive approximations to a fictitious single line spectrum are shown.

§ 1. Introduction

Certain time-distribution functions called ‘ spectra ’ occupy a central position in various theories concerned with the interplay of reversible and irreversible processes. For instance in the theory of the visco-elastic behaviour of high polymers the relaxation and retardation spectra are used; they are time-distribution functions of the elastic modulus and of the compliance respectively. One of these functions describes, in principle, all mechanical properties of the substance to which it refers. In other theories analogous distribution functions referring to thermal and electrical quantities are introduced.

These various ‘ spectra ’ are not functions the form of which can be found directly by the measurement of some quantity under varying conditions. The determination of a spectrum from observational data involves the (numerical) solution of an integral equation. The present paper describes a method for doing this by means of successive approximations. As will be seen, the method is not applicable to all the integral equations which are of interest in the theory of visco-elasticity, but where it does apply it provides a very convenient approach.

The relaxation or the retardation spectrum are suitable for describing the behaviour of real materials under certain physical conditions of linearity and linear superposition. Investigations concerning linear visco-elastic behaviour go back well into the nineteenth century to the work of Maxwell and Boltzmann. This background is reviewed, for instance, by Kolsky (1953). What should be mentioned here is that the phenomenological theory of visco-elasticity is necessarily somewhat extensive and seemingly complicated. There are many observable properties which can serve for investigating the behaviour of a given material experimentally; there are also many possibilities of theoretical description. Of these possibilities the ‘ spectra ’ represent only a special choice. An alternative approach is given by ' memory functions ', which have for instance been used by E. Volterra (1948). The interrelations of the various modes of description are very fully discussed by Gross (1953). A discussion of visco-elastic properties which sets these into perspective with other departures from ordinary elastic behaviour has been given by Taylor (1946).

§ 2. METHODS OF SOLUTION

Many integral equations which are important in the theory of visco-elastic behaviour can be written as convolution-type equations. The general form of such an equation is

$$f(s) = \int K(s-t)g(t)dt. \qquad \ldots\ldots(1)$$

One special form which will be considered in detail later is

$$E''(\omega) = \tfrac{1}{2}\int_{-\infty}^{+\infty} \operatorname{sech}(\tau-\omega)\phi(\tau)\,d\tau. \qquad \ldots\ldots(2)$$

This may be interpreted physically as relating the imaginary part of the dynamic modulus $E''(\omega)$ to the relaxation spectrum $\phi(\tau)$, here a distribution of modulus over a (natural) logarithmic time scale. Thus τ is given by $\tau = \ln(T/T_0)$, where T_0 stands for some standard time and ω is given by $\omega = -\ln(\Omega T_0)$, with Ω the angular frequency.

A general theory exists for the solution of integral equations of type (1) by means of Fourier integrals (Titchmarsh 1937). Reference will be made later to the main results of this theory, and it is convenient to recall them here, very briefly. Under certain conditions, (1) possesses eigenfunctions of type $\exp(-ivs)$ and the associated eigenvalues are given by

$$\frac{1}{\lambda(\nu)} = \int_{-\infty}^{+\infty} K(\xi)\exp(i\nu\xi)\,d\xi. \qquad \ldots\ldots(3)$$

The solution of (1) is therefore, formally, possible by relating the Fourier transform of $f(s)$ to that of $g(t)$ by means of (3), and under certain conditions the solution exists.

When an equation of type (1) is to be solved for the purpose of finding a relaxation spectrum, the function $f(s)$ is given by the empirical dependence of one physical quantity upon another. The most that may then be assumed regarding the ' definition ' of $f(s)$ is that a finite (though perhaps arbitrarily large) number of ordinates $f(s_1)$, $f(s_2)$, ... is available. All these will fall into some finite range of s and will be contaminated with errors. The error-component present in any empirical $f(s)$ entails that, strictly speaking, no convergent solution to the integral equation exists. Any practical procedure for solving the equation under such conditions must therefore provide for the elimination of divergences caused by noise in the data. In addition to this fundamental requirement, it is, of course, desirable that the procedure should be simple.

One possibility is to expand into eigenfunctions, replacing the Fourier integrals by Fourier series for the purpose of numerical work. The convergence of the solution can be enforced by suitable truncation and smoothing. This gives a workable scheme (Roesler and Pearson 1954), but the labour involved is not inconsiderable and some knowledge and practice of Fourier analysis is required for its successful application. It seemed desirable to find, if possible, a simpler, yet systematic approach.†

The idea that an alternative approach could be provided by a method of successive approximations is perhaps rather obvious in view of the wide application which iterative methods have in all types of linear problems. The literature on the iterative solution of linear equations is very extensive (for information and further references see, for instance, Householder 1953) and among the large manifold of schemes proposed for such purposes many would apply to equations of type (1). Many iteration schemes are, furthermore, formulated in general terms such that they include the procedure described below as a special case. Bearing these circumstances in mind, this procedure is put forward as one which has been found useful for the applications with which the authors are concerned. For such applications, the details worked out in this paper are relevant and the authors believe that in this context they are novel.

§ 3. THE ITERATION PROCEDURE

Inserting some trial function or an approximation of order n, $g_{(n)}(t)$, say, into (1), one obtains (in practice by numerical integration):

$$f_{(n)}(s) = \int K(s-t)g_{(n)}(t)\,dt. \qquad \ldots\ldots(4)$$

Using the given $f(s)$ and $f_{(n)}(s)$ defined by (4), it is required to find a correction to $g_{(n)}(t)$ which should bring this nearer to the true $g(t)$. One obvious possibility is to take the correction proportional to $[f(s)-f_{(n)}(s)]$. Thus the next approximation to $g(t)$ would be a $g_{(n+1)}(t)$ defined by

$$g_{(n+1)}(s) - g_{(n)}(s) = p[f(s) - f_{(n)}(s)]. \qquad \ldots\ldots(5)$$

The value of the constant p is disposable and should be fixed so that convergence results.

In order to see whether a procedure based on (5) is feasible, one may consider its effect upon a single eigenfunction of (1). The eigenfunctions are of the form $\exp(-i\nu s)$, and the two functions $f(s)$ and $g(t)$, and all approximations thereto, may be expressed in the form

$$f(s) = \int a(\nu)\exp(-i\nu s)\,d\nu, \quad g(t) = \int b(\nu)\exp(-i\nu t)\,d\nu. \qquad \ldots\ldots(6)$$

With regard to a single eigenfunction component, the iteration is characterized by the successive values of b and of a which it provides, the variable ν now being thought of as fixed. The relevant equations are

$$a_{(n)}(\nu) = b_{(n)}(\nu)/\lambda(\nu), \qquad \ldots\ldots(7)$$

which is true from the definition of an eigenfunction, and

$$b_{(n+1)}(\nu) - b_{(n)}(\nu) = p[a(\nu) - a_{(n)}(\nu)], \qquad \ldots\ldots(8)$$

which is equivalent to (5). In the following, down to equation (11) inclusive, the argument ν will be omitted.

---

† A number of approximation methods for finding relaxation spectra are known, but they involve steps, like the determination of higher derivatives of the empirical functions, which are best avoided in numerical work. A lucid review of such methods has been given by Leaderman (1954).

From (7), equation (8) may be expressed in terms of the b:

$$b_{(n+1)} - b_{(n)} = (p/\lambda)[b - b_{(n)}]. \qquad \ldots\ldots(9)$$

This can be written

$$(b_{(n+1)} - b) - (b_{(n)} - b) = (p/\lambda)[b - b_{(n)}]. \qquad \ldots\ldots(10)$$

By division one obtains

$$\frac{b_{(n+1)} - b}{b_{(n)} - b} = 1 - p/\lambda. \qquad \ldots\ldots(11)$$

If it is arranged that

$$\left|1 - \frac{p}{\lambda(\nu)}\right| < 1 \qquad \ldots\ldots(12)$$

then (11) implies that the relative deviations of the successive approximations $b_{(n)}$ from the true value b decrease according to a geometrical progression.

As the equation (1) is linear and the different eigenfunctions do not interfere, the successful approximation of the functions $g(s)$ and $f(s)$ depends upon fulfilling condition (12) for all possible Fourier components, that is for the λ values given by all values of ν. If the eigenvalues $\lambda(\nu)$ of a certain kernel make it possible to satisfy (12) by giving p some finite value, then iteration according to (5) results in a sequence converging to $g(s)$, assuming that the Fourier transform of $g(s)$ exists. If the eigenvalues are all of the same sign, and if there is a smallest eigenvalue $|\lambda_{\min}| > 0$, then (12) is satisfied if p lies within the range

$$0 < \frac{p}{\lambda_{\min}} < 2. \qquad \ldots\ldots(13)$$

The latitude given by (13) may be used to adjust the convergence to the individual case. If the eigenvalues are not all of the same sign then (12) cannot be satisfied, with one and the same p, for all eigenfunctions, and then an iteration procedure based on (5) cannot be successful.

Since λ is a function of ν the different eigenfunction components of $g(t)$ will not all be approximated with the same speed, even if (12) is satisfied for all. Generally, λ increases with increasing ν, and the higher Fourier components of $g(t)$ will then approach their true values more slowly than the lower ones. In one respect this is desirable, as it implies that the high-frequency noise, due to the random errors in the data, is not unduly amplified. Thus the smoothing, which has to be done explicitly in the Fourier series method, is here automatic. Of course, the automatic smoothing will generally result in some loss of resolving power.

In order to start the iteration procedure, an arbitrary trial function may be used. If the start is made with $g_{(0)} = 0$ and $p = \lambda_{\min}$, then $g_{(1)}$ becomes the usual ' first approximation ' to the spectrum.

§ 4. TREATMENT OF VARIOUS EQUATIONS

4.1. *Given Damping Function*

For the equation (2) the eigenvalues as defined by (3) are

$$\lambda(\nu) = \frac{2}{\pi}\cosh(\nu\pi/2). \qquad \ldots\ldots(14)$$

It is seen that λ is positive for all values of ν, and that $\lambda \geqslant 2/\pi$. The values of p allowed by (13) are $0 < p < 4/\pi$. No useful purpose is served by taking p smaller than $\lambda_{\min}$, so that the practical range is $2/\pi < p < 4/\pi$.

A choice suitable for some applications is to take $p=\lambda_{\min}$ for the first step and then progressively to increase p, say by a factor of $2^{1/4} \simeq 1 \cdot 19$. This implies that at the fifth step one uses $p=2\lambda_{\min}$; thereafter, if more steps are to be made, p is kept constant, as this is the highest value compatible with (13)†. The best start is probably $\phi_{(0)}=0$, $\phi_{(1)}=(2/\pi)E''$. The iteration may be terminated at any stage when satisfactory accuracy appears to have been attained. This and other schemes for choosing p-values are easily worked out on the basis of error estimates for the various Fourier components derivable from (11); examples are given below.

4.2. *Given Dynamic Elasticity*

Another equation of some importance is that which links the relaxation spectrum $\phi(\tau)$ to the real part of the complex modulus, or dynamic elasticity, $E'(\omega)$. This equation is

$$E'(\omega)=\tfrac{1}{2}\int_{-\infty}^{+\infty}[1+\tanh(\tau-\omega)]\phi(\tau)\,d\tau. \qquad \ldots\ldots(15)$$

As the kernel in (15) is not integrable, it is better to consider the differentiated form

$$-\frac{dE'(\omega)}{d\omega}=\tfrac{1}{2}\int_{-\infty}^{+\infty}\operatorname{sech}^2(\tau-\omega)\phi(\tau)\,d\tau. \qquad \ldots\ldots(16)$$

This falls within the scope of the Fourier-integral theory. The eigenvalues are

$$\lambda(\nu)=\frac{2}{\nu\pi}\sinh(\nu\pi/2). \qquad \ldots\ldots(17)$$

All values of λ are positive and $\lambda_{\min}=1$. Thus the iteration method is applicable to (16). One numerical differentiation of the $E'(\omega)$ data is necessary to find the function on the left-hand side in (16). The range of admissible p-values is $0<p<2$ in principle, and the useful range is $1<p<2$.

4.3. *Given Stress Relaxation Function*

A third equation of interest is the one which links the relaxation spectrum to the stress relaxation function (or the retardation spectrum to the creep-recovery function). This is

$$\sigma(\tau)=\int_{-\infty}^{+\infty}\phi(\omega)\exp\left[-\exp(\tau-\omega)\right]d\omega. \qquad \ldots\ldots(18)$$

As in the case of equation (15), it is preferable to consider the differentiated form

$$-\frac{d\sigma(\tau)}{d\tau}=\int_{-\infty}^{+\infty}\phi(\omega)\exp\left[\tau-\omega-\exp(\tau-\omega)\right]d\omega. \qquad \ldots\ldots(19)$$

The kernel of (19) is

$$K(\tau-\omega)=K(\xi)=\exp\left[\xi-\exp\xi\right]. \qquad \ldots\ldots(20)$$

This function is continuous and integrable, and possesses a continuous derivative. Accordingly its Fourier transform is well defined, though it is not readily expressed by known functions. However, its behaviour can be investigated numerically. This has been done and the detailed results are given elsewhere (Roesler 1955).

---

† In principle, once it may be assumed that the lower Fourier components have attained their true values, it would be possible to use even larger values of p. However, certain precautions would then be necessary. It seems most practicable to keep within (13).

As the kernel (20) is neither symmetric, nor antisymmetric, its eigenvalues are complex. (In physical terms, the kernel produces both a change of amplitude and a phase shift.) The numerical investigation shows that both the real and the imaginary part of the Fourier transform of the kernel (20) oscillate in sign. From this it must be concluded that equation (19) cannot be solved by an iteration based on the rule (5), though it may be solved by means of Fourier series.

§ 5. EXAMPLES

Data concerning the visco-elastic behaviour of polyisobutylene in shear, due to a co-operative research project, have been published by Marvin (1953 a, b). The shape of the strongest peak in the spectrum, given by the rubber–glass transition, has been determined from these data by use of the Fourier series method (Roesler and Pearson 1954). In other regions, the amplitude of the spectrum is very much smaller, but its form there is nevertheless of considerable physical interest. The determination of the spectrum over the whole range of data is in such a case necessarily laborious. The enormous span of amplitudes, amounting to a factor of 10^6, makes the division of the work into stages inevitable. When using the Fourier series method, one has to apply it several times, starting by the evaluation of the main peak and subtracting at each step the effect of the partial spectrum already found from the data. The method then becomes one of iteration, similar to that proposed here, but in which the successive corrections are found by means of Fourier series instead of by means of (5). Such a procedure has been tried and was found to be workable. The simpler iteration method based on (5) is, however, more efficient for the type of problem under consideration and it is possible by this means to deal from the start with the whole range of data, despite the very large variation in the damping. Another advantage of the iteration based on (5) is that the significance of the remainders $G''-G''_{(n)}$ need not be considered at each step.

A very large range of amplitudes causes difficulties in the choice of p-values. There is then no way of establishing, before the actual computation is done, whether a given sequence of p-values which increase from $p=\lambda_{\min}$ to $p=2\lambda_{\min}$ will not in later stages, when $p=2\lambda_{\min}$ is used, produce oscillations in the regions of small spectral amplitude. Such oscillations may even arise if the largest p used is less than $2\lambda_{\min}$, and though they should ultimately die out they make it impossible to terminate the whole computation at an arbitrary step. One way of overcoming this difficulty is never to raise p above $\lambda_{\min}$. The convergence of the whole process is then slow, but is systematic over the whole range. Another possibility is to step up the p-values to $p=2\lambda_{\min}$ fairly quickly, thus obtaining fast convergence at the main peak, and ultimately, when the shape of the main peak has ceased to change substantially, to reduce p again to $p=\lambda_{\min}$.

Such a treatment of the Marvin data (1953 a, addendum) for $G''(\omega)$ of polyisobutylene has been carried out by the authors. Eight iterations were done, the values of p, in multiples of $\lambda_{\min}=2/\pi$, being 1, 1·19, 1·41, 1·68, 2, 2, 1, 1. The treatment and its results are summarized in the table. The first column shows the kernel. The straight values of $\frac{1}{2}\operatorname{sech}(\tau-\omega)$ must be corrected for the number of ordinates per unit of $\tau-\omega$; in Marvin's tabulation, the values of G'' are spaced at 0·4 decade, which is $0{\cdot}922\tau$ units. This spacing of ordinates was accordingly used. If one is free to choose, it is advisable to use somewhat narrower intervals, of order 0·1 . . . 0·2 decade, and the higher the accuracy of the data, the smaller the interval that should be taken. For the computation, the values of the kernel as given in column 1 are put on a strip, which can be laid along the column of ϕ-values. Each value of G'' can then be obtained by accumulating the products of

Table. Evaluation of the Relaxation Spectrum for Shear of Polyisobutylene

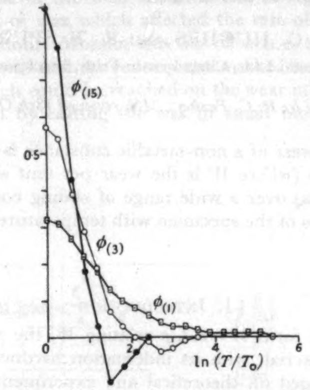
Values of the kernel of equation (2), for an interval of 0·4 in log T	Log T	$\phi_{(1)}=(2/\pi)\,G''$ (in units of dyn cm^{-2} throughout)	$\phi_{(8)}$	$\phi_{(\text{Fourier})}$	
	$-6{\cdot}0$	$3{\cdot}93\times10^{4}$	$1{\cdot}79\times10^{4}$		
	$-5{\cdot}6$	7·62 ,, ,,	7·38 ,, ,,		
	$-5{\cdot}2$	$1{\cdot}18\times10^{5}$	9·60 ,, ,,		
	$-4{\cdot}8$	1·69 ,, ,,	$1{\cdot}68\times10^{5}$		
	$-4{\cdot}4$	2·18 ,, ,,	2·59 ,, ,,		
$0{\cdot}9201\times10^{-6}$	$-4{\cdot}0$	2·45 ,, ,,	2·67 ,, ,,		
$0{\cdot}2324\times10^{-5}$	$-3{\cdot}6$	2·58 ,, ,,	2·71 ,, ,,		
$0{\cdot}5814\times10^{-5}$	$-3{\cdot}2$	2·59 ,, ,,	2·74 ,, ,,		
$0{\cdot}1460\times10^{-4}$	$-2{\cdot}8$	2·50 ,, ,,	2·80 ,, ,,		(*)
$0{\cdot}3668\times10^{-4}$	$-2{\cdot}4$	2·24 ,, ,,	2·02 ,, ,,		
$0{\cdot}9210\times10^{-4}$	$-2{\cdot}0$	2·07 ,, ,,	2·16 ,, ,,		
$0{\cdot}2313\times10^{-3}$	$-1{\cdot}6$	1·84 ,, ,,	1·97 ,, ,,		
$0{\cdot}5813\times10^{-3}$	$-1{\cdot}2$	1·56 ,, ,,	1·33 ,, ,,		
$0{\cdot}1460\times10^{-2}$	$-0{\cdot}8$	1·42 ,, ,,	1·53 ,, ,,		
$0{\cdot}3668\times10^{-2}$	$-0{\cdot}4$	1·28 ,, ,,	0·98 ,, ,,		
$0{\cdot}9210\times10^{-2}$	$0{\cdot}0$	1·32 ,, ,,	1·52 ,, ,,		
$0{\cdot}2313\times10^{-1}$	$0{\cdot}4$	1·41 ,, ,,	1·15 ,, ,,		
$0{\cdot}5806\times10^{-1}$	$0{\cdot}8$	1·75 ,, ,,	1·22 ,, ,,		
0·1427	$1{\cdot}2$	2·42 ,, ,,	1·68 ,, ,,		
0·3165	$1{\cdot}6$	3·87 ,, ,,	1·01 ,, ,,		
0·4605	$2{\cdot}0$	7·14 ,, ,,	1·31 ,, ,,		
0·3165	$2{\cdot}4$	$1{\cdot}39\times10^{6}$	6·92 ,, ,,		
0·1427	$2{\cdot}8$	2·62 ,, ,,	$1{\cdot}60\times10^{6}$		
$0{\cdot}5806\times10^{-1}$	$3{\cdot}2$	4·83 ,, ,,	2·42 ,, ,,		
$0{\cdot}2313\times10^{-1}$	$3{\cdot}6$	9·00 ,, ,,	4·12 ,, ,,		
$0{\cdot}9210\times10^{-2}$	$4{\cdot}0$	$1{\cdot}69\times10^{7}$	$1{\cdot}00\times10^{7}$		(*)
$0{\cdot}3668\times10^{-2}$	$4{\cdot}4$	3·12 ,, ,,	$8{\cdot}18\times10^{6}$		
$0{\cdot}1460\times10^{-2}$	$4{\cdot}8$	6·08 ,, ,,	$3{\cdot}74\times10^{7}$		
$0{\cdot}5813\times10^{-3}$	$5{\cdot}2$	$1{\cdot}12\times10^{8}$	4·00 ,, ,,		
$0{\cdot}2313\times10^{-3}$	$5{\cdot}6$	2·16 ,, ,,	9·31 ,, ,,		
$0{\cdot}9210\times10^{-4}$	$6{\cdot}0$	4·11 ,, ,,	$2{\cdot}20\times10^{8}$		
$0{\cdot}3668\times10^{-4}$	$6{\cdot}4$	7·48 ,, ,,	4·59 ,, ,,	$5{\cdot}20\times10^{8}$	
$0{\cdot}1460\times10^{-4}$	$6{\cdot}8$	$1{\cdot}27\times10^{9}$	8·80 ,, ,,	8·50 ,, ,,	
$0{\cdot}5814\times10^{-5}$	$7{\cdot}2$	1·97 ,, ,,	$1{\cdot}72\times10^{9}$	$1{\cdot}75\times10^{9}$	
$0{\cdot}2324\times10^{-5}$	$7{\cdot}6$	2·68 ,, ,,	3·68 ,, ,,	3·70 ,, ,,	
$0{\cdot}9201\times10^{-6}$	$8{\cdot}0$	2·84 ,, ,,	4·64 ,, ,,	4·49 ,, ,,	
	$8{\cdot}4$	2·11 ,, ,,	2·29 ,, ,,	2·54 ,, ,,	
	$8{\cdot}8$	1·20 ,, ,,	$5{\cdot}04\times10^{8}$	$4{\cdot}10\times10^{8}$	
	$9{\cdot}2$	$5{\cdot}73\times10^{8}$	$9{\cdot}66\times10^{7}$		

The first column shows the appropriately scaled values of the kernel of the integral equation, in the position for evaluating G'' for log $T=2{\cdot}0$. The second shows the decadic logarithm of the characteristic time in seconds. The third shows the first approximation to the spectrum given by the scaled-down damping function, using the data published by Marvin (1953 a). The fourth column shows the eighth approximation, which can be considered as final with reference to the accuracy of these data. The fifth shows, for comparison, the spectral amplitudes obtained by Roesler and Pearson (1954) by means of Fourier series, from the same data. The asterisk (*) indicates oscillations (see text).

adjacent numbers when the central value of the strip is opposite the appropriate value of ϕ. The correction to each value of ϕ is found, from (5), by subtracting the corresponding value of $G''_{(n)}$ from the true G'' (the data) and multiplying the difference by p. The first approximation $\phi_{(1)}$ is obtained by multiplying the G'' values by $2/\pi$. The third column of the table shows this first approximation. The fourth column shows the final, eighth approximation. At this stage, all corrections to the spectrum, everywhere, became so small that further work could not result in significant changes. For comparison, the results obtained previously by means of Fourier series, in the region where the single Fourier analysis can be expected to be reliable, are shown in the next column of the table. The agreement is obviously good.

A number of values of ϕ have been marked with an asterisk (*). It is seen, most easily by making a plot of $\phi_{(8)}$ for this region, that the spectrum here shows short oscillations. It would seem very unlikely that these oscillations are physically real, and one might consider it justifiable to remove them by graphical smoothing. This was not, however, done in this paper, since the oscillations are not due to errors in the approximation process. They are mathematically proper consequences of a slight roughness of the E'' data in this region, which can also be noticed on a large-scale non-logarithmic plot. If the ordinates were more narrowly spaced, such oscillations, provided they are due to random errors in the data, would be reduced.

The speed of the convergence in a given case depends not only on the choice of p-values, but also on the form of the true spectrum. The sharper the peaks in the true spectrum, the more slowly do the results of the successive iterations approach it. The sharpest peak that can occur is a single line spectrum. Convergence in the proper sense is then impossible, as the δ-function cannot be represented by a series. The damping function corresponding to a single line spectrum, with the line at $\omega = 0$, is given by $\frac{1}{2}$ sech ω. In order to gain some experience of how the iteration process works in the case of sharp peaks, the authors have applied it to this function. Such a treatment should not be done analytically (as, in principle, it could be when the damping function is given analytically), but numerically, so that the rounding-off errors make it realistic. Accordingly, $\frac{1}{2}$ sech ω was considered as a numerically given function, using only two figures. The ordinates were spaced at intervals of 0·3 units of ω. There are 35 non-vanishing ordinates and the ordinate sum of a correctly normalized spectrum is 3·33. The rounding-off error for each ordinate is 0·01 and the probable error for the sum $0{\cdot}01\sqrt{35} = 0{\cdot}06$, which is 2%. The normalization of the spectrum is given by the lowest Fourier components which are thus likely to drift by about 2% at each step of the computation, with the total drift increasing as the root of the number of iterations. This drift will, of course, be counteracted by the iteration itself so long as p is less than $2\lambda_{\min}$. Accordingly, if it is desired that the normalization should stay correct within certain limits throughout the iteration, the values of p should not be raised above a figure which is smaller, by a suitable margin, than $2\lambda_{\min}$. The theoretically optimum choice of p under such conditions presents a difficult and interesting problem, which will not be considered here. In practice, fixing p by a guess is unlikely to lead to much loss in convergence. Should oscillations arise from p having been chosen too large, the mistake can be rectified during the computation. For the evaluation of the line spectrum discussed here, the values of p used, in multiples of $\lambda_{\min}$, were 1, 1·3, then constantly 1·6. The normalization was checked at each step and it was found that it did not show oscillations larger than ± 3% of the correct value. After fifteen steps the difference $E'' - E''_{(15)}$ was smaller than 0·01 at any point and the accuracy of the presupposed data could therefore be considered as exhausted. The figure shows a plot of the first approximation $\phi_{(1)} = (2/\pi)\,E''$, of the third approximation $\phi_{(3)}$ and of the last approximation $\phi_{(15)}$. This example shows that comparatively rough data will support many iterations, provided they contain no systematic errors.



Successive approximations to a sharp line in the spectrum.
Line centre at $T = T_0$, $\ln(T/T_0) = 0$.

Acknowledgments

The authors are grateful to Dr. H. Kolsky and Mr. J. J. Benbow for criticism, advice and discussion.

References

Gross, B., 1953, *Mathematical Structure of the Theories of Visco-elasticity* (Paris: Hermann).
Householder, A. S., 1953, Principles of Numerical Analysis (New York–Toronto–London: McGraw-Hill).
Kolsky, H., 1953, *Stress Waves in Solids* (Oxford: Clarendon Press).
Leaderman, H., 1954, *J. Appl. Phys.*, **25**, 294.
Marvin, R. S., 1953 a, *National Bureau of Standards*, Report No. 2305 and addendum; 1953 b, *Proc. 2nd Int. Congr. on Rheology, Oxford*, p. 156.
Roesler, F. C., 1955, *Proc. Phys. Soc.* B, **68**, 89.
Roesler, F. C., and Pearson, J. R. A., 1954, *Proc. Phys. Soc.* B, **67**, 338.
Taylor, G. I., 1946, *J. Instn Civ. Engrs*, **26**, 486 (James Forrest Lecture).
Titchmarsh, E. C., 1937, *Introduction to the Theory of Fourier Integrals* (Oxford: Clarendon Press).
Volterra, E., 1948, *Riv. Nuovo Cimento*, **4**, 1.

An Experimental Investigation of Non-Metallic Wear

By G. HUGHES and R. T. SPURR
Ferodo Ltd., Chapel-en-le-Frith, Stockport

Communicated by R. C. Parker; MS. received 18th October 1954

Abstract. **The rate of wear of a non-metallic substance is shown to be given by the equation $W=KL/p$ (where W is the wear per unit sliding distance, L the load and p the hardness) over a wide range of sliding conditions provided the variation of the hardness of the specimen with temperature is taken into account.**

---

§ 1. Introduction

Equations of the form $W=KL/p$ relating W, the wear per unit sliding distance of a material, with its indentation hardness p, and the load L, have been proposed on theoretical and experimental grounds by Holm (1946), Burwell and Strang (1952), Archard (1953) and Rabinowicz (1953). The wear equation involves the hardness p because it is generally assumed that, though true contact occurs between two surfaces at small isolated areas, the total true area of contact for plastic materials is related to L/p, and that the wear process involves the formation of wear particles at the isolated contact areas. Archard considers K to be related to the probability of a wear particle being formed at a particular contact area. Experimental verification of the equation has previously been limited to showing that the rate of wear is proportional to the load and there has been no direct experimental verification of the inverse relationship between wear and hardness. The work described below was undertaken to investigate the relationship between wear and hardness directly, by sliding a low melting point material (a wax) against a cast iron disc and measuring the rate of wear and the temperature reached at the rubbing surface over a range of loads and speeds. The variation with temperature of the hardness of the wax was then measured and the appropriate values of W, L and p substituted in the equation.

§ 2. Experimental Details

2.1. *Measurement of Wear*

Figure 1 shows the apparatus used for measuring wear and friction. The cast iron plate A, 35 cm in diameter, was connected to a 3 h.p. electric motor providing running speeds ranging from 60 to 300 r.p.m. (i.e. linear sliding speeds from 100 to 500 cm sec^{-1}). A torque arm B was freely suspended above the plate, and restrained at one end by a friction-measuring pendulum device, the guide C ensuring that the restraining cord acted at a constant radius. At the other end of the torque arm, the specimen holder D was freely hinged at E, so that dead weight loading could be applied.

The surface of the plate was prepared by abrading it with a carborundum stone lubricated with paraffin oil while the plate was running at high speed, and then cleaning the surface with cotton wool soaked in carbon tetrachloride. This procedure was repeated before each wear measurement, for the wax tended to smear over the disc of the wear machine and at higher loads and speeds to build up a thick film of wax which affected the rate of wear, and also reduced the coefficient of friction. Montan wax was chosen as the test material because of its low melting point and its consequent marked softening at the relatively low temperatures which could be reached on the wear machine. The specimens, which were prepared by casting the wax in small moulds, had rubbing areas of ½ in. by ½ in.

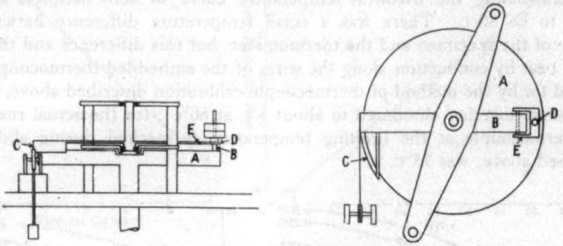


Figure 1. Wear machine.

The rate of wear per unit sliding distance was measured over a range of loads from 845 to 3000 grammes and measurements were made at four different speeds for each load. The wear was determined by measuring the loss in weight after 5-minute runs, for it was found that wear increased linearly with sliding distance, and the average of six such runs taken. The wear results were quite reproducible, except perhaps at very low rates of wear when slight contamination of the wear machine disc could affect results.

2.2 *Measurement of Surface Temperature*

A wax specimen was cast with a thermocouple incorporated in the mould in such a way that the junction was close to the surface to be used for sliding. When the wax specimen was run on the wear machine the thermocouple registered room temperature until the junction reached the surface when the temperature increased sharply to a value characteristic of the sliding speed and load. Thermocouple readings were recorded at various loads and speeds. It was observed that no matter how severe the sliding conditions the thermocouple reading would not rise above a certain value. The thermocouple was therefore calibrated by assuming that this limiting value corresponded to the melting point of the wax.

2.3 *Measurement of Variation of Hardness with Temperature*

The hardness tester used consisted of a metal bar 24 in. long pivoted at the centre. Halfway along one arm the diamond indenter from a Vickers hardness tester was affixed and counter-balance weights hung on the other arm until the bar just balanced in the horizontal position. The load required for making the indentations was attached to the end of the part holding the diamond indenter.

Underneath the indenter a microscope heating stage was placed and covered with a lid made of a very poor conductor with a hole in the centre just large enough to allow the entrance of the diamond indenter.

Preliminary measurements showed that the hardness of the wax was constant over the experimental range 10–3000 g load. Before making any indentations at elevated temperatures the apparatus was maintained at the required temperature for ten minutes to ensure that thermal equilibrium was reached. A fresh sample of wax was used at each temperature and indentations were made at temperature intervals of 5°c up to a maximum of 75°c according to the microscope heating stage thermometer. As at higher temperatures the wax did not give a clear-cut indentation and creep occurred, the melting point was obtained by extrapolating the hardness–temperature curve to zero hardness and was found to be 80°c. There was a small temperature difference between the surface of the specimen and the thermometer, but this difference and the slight loss of heat by conduction along the wires of the embedded thermocouple were allowed for by the method of thermocouple calibration described above. These two factors together amounted to about 5°c at 80°c; for the actual reading of the thermocouple at the limiting temperature it reached during sliding, as described above, was 75°c.

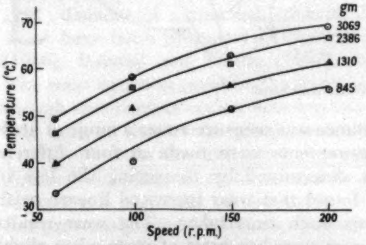


Figure 2. Variation of temperature with speed at different loads.

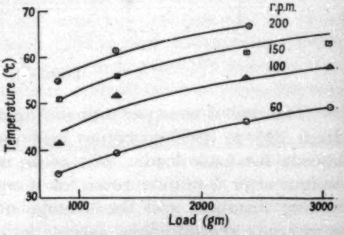


Figure 3. Variation of temperature with load at different speeds.

§ 3. RESULTS

Thermocouple readings obtained at various speeds and loads and calibrated as described above are plotted in figures 2 and 3, and smooth parallel curves drawn through the points. These curves were drawn so that the variation of temperature with speed and variation of temperature with load gave sets of curves which were mutually consistent and lay as close as possible to the experimental points. The variation of the hardness of wax with temperature is shown in figure 4.

The wear was measured at fourteen different sets of sliding conditions. The hardness of the wax was obtained for each set of sliding conditions from figures 2, 3 and 4 and the wear figures and hardnesses used to plot figure 5 where W/L is shown plotted against $1/p$. It can be seen that the points fall about a straight line showing that the rate of wear is indeed given by the equation $W=KL/p$. The value of K can be obtained from the slope of the line and using this value of K the wear–load curves for the four sliding speeds were calculated and are shown in figure 6. The actual experimental wear figures are shown by circles.

§ 4. Friction Effects

The softening of the wax with increasing temperature suggests that the true area of contact between specimen and plate, and therefore the coefficient of friction μ, might increase with temperature. This however was not the case for μ decreased from 0·4 at room temperature to about 0·30 at 65°c. King and Tabor (1953) measured the variation of the coefficient of friction μ, shear strength s, and hardness p, with temperature of a number of plastics and found that the equation

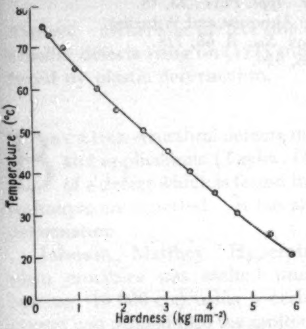


Figure 4. Variation of hardness of the wax with temperature.

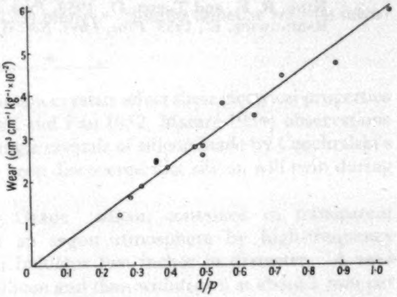


Figure 5. Wear plotted against $1/p$.

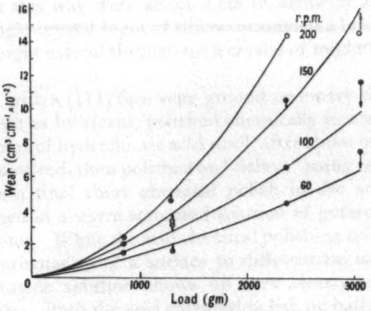


Figure 6. Variation of wear with load at different speeds. Experimental wear values shown by circles.

$\mu = Cs/p$, held over a wide range of temperatures; C for polythene was 1·1, and for Perspex 2·4. The shear strength of wax cylinders was therefore measured at temperatures from 20 to 70°c. There was considerable scatter in the measured shear strength, but the value of s/p was found to be more or less constant (0·05 to 0·08) over the whole temperature range, though considerably less than μ. The difference between μ and s/p was probably due to the wax failing in a brittle manner in the shear strength measurements whereas a ductile type of shear occurred at the contact areas during sliding.

ACKNOWLEDGMENTS

The authors wish to express their thanks to members of Ferodo Technical Division for criticisms and suggestions, and to Dr. R. C. Parker, Technical Director of Messrs. Ferodo Ltd. for permission to publish this paper.

REFERENCES

ARCHARD, J. F., 1953, *J. Appl. Phys.*, **24,** 981.
BURWELL, J. T., and STRANG, C. D., 1952, *J. Appl. Phys.*, **23,** 18.
HOLM, R., 1946, *Electric Contacts* (Uppsala : Almquist and Wiksells).
KING, R. F., and TABOR, D., 1953, *Proc. Phys. Soc.* B, **66,** 728.
RABINOWICZ, E., 1953, *Proc. Phys. Soc.* B, **66,** 929.

Lamellar Defects in Single Crystals of Silicon

By J. FRANKS, G. A. GEACH and A. T. CHURCHMAN
Research Laboratory, Associated Electrical Industries Ltd., Aldermaston Court, Aldermaston, Berks.

MS. received 9th November 1954

Abstract. Otherwise perfect single crystals of silicon have been shown to contain lamellar defects lying on (111) and (123) planes. Similar lamellae may be introduced by plastic deformation.

Because structural defects in silicon crystals affect their electrical properties and applications (Taylor, Odell and Fan 1952, Mataré 1954) observations of a defect which is found in single crystals of silicon made by Czochralski's technique are reported. It has also been discovered that silicon will twin during deformation.

Johnson, Matthey 'Hyperpure Grade' silicon contained in transparent silica crucibles was melted under an argon atmosphere by high-frequency heating (10 000 c/s) using a carbon inductor two inches in diameter. A seed crystal was dipped into the molten silicon and then withdrawn at about 1 mm per minute: it was kept rotating at about 50 rev min^{-1} throughout a run. An essential condition to avoid polycrystallinity was that before withdrawal was commenced the seed be completely wetted wherever in contact with liquid. Crystals prepared in this way were about 2 cm in diameter and weighed 15 to 20 gm. In every single-crystal ingot of silicon examined a few very fine lamellae were found; these might extend throughout a crystal or might begin or terminate during its growth.

Crystal slices cut with a (111) face were ground on emery cloths to 320 grade, using methylated spirit as lubricant, polished chemically in a mixture of 10 parts of nitric acid with 1 part of hydrofluoric acid until, after about one hour, the emery scratches had been removed, then polished on 'Selvyt' using magnesia or alumina in water, and given a final short chemical polish in the acid mixture. The specimens were etched in a warm saturated solution of potassium hydroxide for times up to two minutes. While the acid chemical polishing solution is sufficiently sensitive to crystal orientation of a surface to differentiate large twinned areas, the potassium hydroxide solution shows up very clearly small imperfections present in the crystals. Both the acid chemical polishing bath and the potassium hydroxide etch attack strained areas in the silicon more than surrounding strain-free material.

Measurements made of the traces of these lamellae on two faces whose crystallographic orientations had been determined by x-ray Laue method showed that the lamellae were parallel to (111) planes. In this way they resembled twins in silicon. Their width, however, was often unusually small for twins in crystals of high symmetry. Examination under the electron microscope using carbon replicas (Bradley 1954) showed that they might be as little as 100 Å thick (figure 1, Plate). The lamellae tended to occur less frequently on that (111) plane most nearly normal to the growth direction, than on other {111} planes.

In crystals grown in the [111] direction, no lamellae were observed on the (111) plane parallel to the liquid–solid interface.

Back reflection x-ray examination, using a microfocus tube so arranged that a spot of the specimen 1 mm in diameter was examined and that the Laue spot was an image of the irradiated spot, showed that there was considerable strain in the silicon about the lamellae. The Laue spots were crossed by dark lines parallel to the traces in the specimen surface of the lamellae. Twin material should have left an unblackened line on the spot and it appears that these dark lines are due to a disturbance in the matrix adjacent to the twins. From the width of these black traces this strained region was found to be between ten and fifty times as thick as the lamellae. A broad trench whose width was about that of this strained band was etched in the material about the lamellae and could be observed using an electron microscope. This trench had a flat bottom parallel to the original (111) surface being examined, and was bounded by cliff-like edges parallel to the lamellae.

Hexagonal etch pits were formed on {111} faces by the potassium hydroxide etchant. These pits only occurred along the lamellae : they were always observed at points where a change in the breadth occurred in the lamellae and at the ends, that is at points where localized strain is to be expected in the crystal (figure 2, Plate).

A small number of traces which could not be those of {111} planes has been observed. These were possibily due to similar lamellae parallel to {123} planes in the matrix crystal.

The evidence that silicon will twin during deformation was obtained from experiments in which small bars were bent by three-point loading. Specimens 3 mm square with a span of 10 mm were heated in the h.f. furnace and bent under loads of 1800 to 4500 gm at about 1300°c. They were polished and examined before and after bending. It was found that the deformation had produced a small number of twins having the form of the lamellae just described.

Of possible boundary planes between a twin and a completely surrounding matrix two parallel (111) planes may be completely coherent. No other part of the twin surface can be coherent. This is in accordance with the plate-like form of the lamellae observed in the (111) planes. Very commonly the lamellae reached the surface of the ingot : it is possible thus to avoid all incoherent silicon to silicon interfaces, except where changes in width occurred. It is possible that a completely enclosed twin may be bounded, at least in part, by faces other than (111) but on which the parameter of the super lattice of coincidence is not many times greater than the lattice parameter. Such planes must, of course, have low index zone axes, such as [111], and the (123) plane is a plane of this nature.

ACKNOWLEDGMENTS

The micro-focus back reflection x-ray experiment was made by Mr. P. J. Holmes of this Laboratory and his help is acknowledged.

The authors wish to thank Dr. T. E. Allibone for permission to publish this paper.

REFERENCES

BRADLEY, D. E., 1954, *Brit. J. Appl. Phys.*, **5**, 96.
MATARÉ, H. F., 1954, *Z. Naturf.*, **9a**, 698.
TAYLOR, W. E., ODELL, N. H., and FAN, H. Y., 1952, *Phys. Rev.*, **88**, 867.

RESEARCH NOTES

The Barrier Height of Point-Contact Germanium Diodes inferred from Measurements of the Voltage Dependence of Capacitance

By F. F. ROBERTS and J. R. TILLMAN
Post Office Engineering Research Station, Dollis Hill, London, N.W.2

MS. received 14*th October* 1954

The capacitance C of a point-contact germanium diode, subjected to a reverse bias of voltage V, has been measured with apparatus utilizing the basic circuit of most commercial Q-meters, but having further attention paid to stability of frequency, to resolving power of the variable capacitor and, because the signal applied to C had sometimes to be less than 50 mv, to the sensitivity. Spreading resistance was unimportant in all measurements.

Results appropriate to the whisker–germanium contact were obtained at the cost of destroying the unit, by measuring the residual capacitance of the body and lead wires after the whisker had been forcibly removed from, and arranged to remain a small distance off, the germanium. The capacitance of a complete unit mounted conventionally was, depending on its construction, anything up to 0·5 pF greater than the values now given.

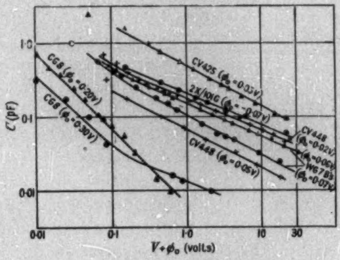


Figure 1. The voltage dependence of the capacitances of some point-contact germanium diodes.

Figure 1 shows results for several types obtained with a signal frequency of 1·2 Mc/s ; separate tests indicated that the results were independent of the signal frequency. The presentation, log C against log $(V+\phi_0)$, gives equal prominence to data for V greater than and less than 1 v and to large and small values of C. ϕ_0 is a constant for each unit, chosen to give each curve a close approach to a straight line. The reciprocal of the slope represents the index in the relationship $C^{-n} \propto (V+\phi_0)$; n is seen to be approximately 2 (rather than 1 or 3), for all but the CG8's. The curves refer to individual units, but others of the same types have shown similar variations of C with V even though C, at a given voltage, differed by a factor of up to 3. The curves must not be extrapolated to the left.

Figure 2 shows results for three units type GEX64. The points for $(V+\phi_0) > 0{\cdot}1$ v refer to a test frequency of 10 Mc/s (known to be unimportant), while those for $V=0$ refer to frequencies up to 0·5 Mc/s only. The slopes of the curves are closely $-1/3$, suggesting $n=3$. When the extreme left-hand point of each curve is neglected, as perhaps it should be because $V < 0{\cdot}025$ v $(=\boldsymbol{k}T/q)$, $|\phi_0|$ remains very small $(< 0{\cdot}025$ v).

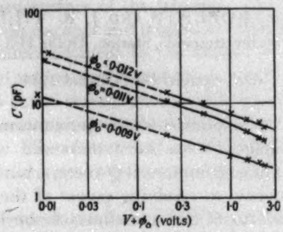


Figure 2. The voltage dependence of the capacitances of three diodes type GEX64.

The simplest interpretation of the value $n=2$ is that the net donor concentration in the germanium is constant throughout the region (of radial thickness extending to a few microns) swept by the space charge barrier layer for the range of bias voltages used. It implies negligible diffusion of acceptor impurities from the catswhisker (or from the possibly contaminated surface of the germanium) during the electrical forming of the contact. The value $n=3$ on the other hand suggests a uniform gradient of net donor concentration through the barrier layer (thinner since able to withstand less reverse bias) of the GEX64 type diodes, a gradient that can only have been set up by the diffusion of acceptor impurities from the catswhisker to distances of about a micron into the germanium during the electrical forming.

The voltage intercept ϕ_0 is related to the spontaneous thickness of the barrier layer in the absence of any applied bias. Schwarz and Walsh (1953) have shown that, when the experimental results are consistent with $n=2$ over a wide voltage range, as appears for several of the diodes above, the value of ϕ_0 should be only about 0·025 v less than the barrier height for electrons approaching the metal from the semiconductor. The donor concentrations of, and hence the Fermi levels in, the material used for the diodes of figure 1 can be inferred from their known turnover voltages, ranging from 30 to 120 volts (by comparison with results obtained for laboratory-made diodes using monocrystalline germanium of known resistivity). It then follows that the barrier height for electrons approaching the semiconductor from the metal is, at least for all these diodes except the CG8's, less than or barely equal to half the forbidden energy gap for germanium.

Other measurements at these laboratories indicate that the electrical forming of tungsten contacts on monocrystalline germanium, to an extent suitable for the production of good diode characteristics, does not reduce the hole-injection ratio (at 1 mA forward current) to less than 0·2 for germanium of resistivity 1–5 Ωcm. One purpose of the note therefore is to draw attention to the discrepancy between the small barrier heights obtained from the capacitance measurements and those (much greater) that may be inferred from the measurements of injection ratio (Hogarth 1953, Gunn 1954, Swanson 1954).

ACKNOWLEDGMENT

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in the note.

REFERENCES

GUNN, J. B., 1954, *Proc. Phys. Soc.* B, **67,** 575.
HOGARTH, C. A., 1953, *Proc. Phys. Soc.* B, **66,** 845.
SCHWARZ, R. F., and WALSH, J. F., 1953, *Proc. Inst. Radio Engrs, N.Y.*, **41,** 1715.
SWANSON, J. A., 1954, *J. Appl. Phys.*, **25,** 314.

REVIEWS OF BOOKS

Einführung in die Physik, Band 3, *Optik und Atomphysik*, by R. W. Pohl. Pp. viii + 356. (Berlin, Göttingen, Heidelberg: Springer.) DM 29.70.

The author describes a large number of experiments, many of them lecture demonstrations, and then gives a very brief account of the associated theories. In this way, starting from rudimentary beginnings, he is able to introduce a large number of topics very rapidly. Near the beginning of chapter 2, the reader is introduced to the laws of reflection and of refraction. On page 12, a lens is considered as an assembly of prisms and on page 17 there is an elementary treatment of resolving power. On page 19 the idea of dispersion is first introduced and on page 21 the properties of infra-red and ultra-violet radiation are described. The chapter ends with a discussion of mirror systems (pentagonal prisms, corner cubes, etc.). Chapter 4 begins on page 32 with nodal points and principal planes and includes the third-order aberrations, the Schmidt camera, achromatization, telescopic systems, magnification, field of view, stops, and the theory of perspective. The lectures on which this book is based are probably useful to the students for whom they are intended especially when illustrated by many good lecture demonstrations (some of which are described) and illuminated by the personality of the distinguished author. As the above summaries suggest they are so different from a course of lectures in a British University that a translation could not be recommended to British students.

One is shocked by the ease with which topics appear and then are allowed to fade out so that one chapter includes work which we should regard as school work and also material suitable for a Special Honours course. Yet, for the teacher of Physics, it is not sufficient to review this book as the Victorian lady reviewed *Antony and Cleopatra* with the phrase " How unlike the home life of our own dear Queen ". We need not imitate, but can it be that the somewhat pedestrian systematic, logical exposition (" starting at the beginning and going on to the end "), which constitutes our typical lecture course, would be improved by an occasional ' preview ' of something interesting which logically should come much later, and that students should occasionally be taken out of their depths, even though most of them prefer to paddle! Any British University lecturer in optics who has given the same lectures for five years might profitably read this book and reconsider his course before saying " Same again ".

Part II on ' Atomic Physics ' includes sections on atomic and molecular spectra, temperature radiation, fundamental particles and on the author's own special subject ' Quantum Optics of Solids '. The treatment of spectra is in terms of the Bohr orbits. Here, and in a number of other places, a more modern viewpoint could, with advantage, be adopted. The 565 illustrations include many perspective drawings of lecture bench apparatus where line diagrams would be more useful. There are also a large number of excellent photographic illustrations.

R. W. DITCHBURN.

Der Ultraschall und seine Anwendung in Wissenschaft und Technik, 6th Edn, by L. Bergmann. Pp. xvi + 1114. (Stuttgart: Hirzel, 1954.)

The science of ultrasonics has travelled far since Dr. Bergmann published the first edition of his well-known book in 1937, how far can be succinctly seen by comparing the expansion in size, 250 to 1100 pages. Over the years it has been the constant reference book of workers in this field. Here they will find a complete up-to-date bibliography of pure and applied ultrasonics, in which every paper published finds its due place at the hands of this industrious abstractor.

In some respects the warp and woof of six editions shows through the material. It is questionable whether diagrams and photographs of apparatus now only of historic interest should continue to be reproduced, or the latest and perhaps more conclusive researches appear as small type additions to earlier work. One would like as much space given to English and American industrial applications as to German ones even if data about the latter are nearer to the author's hand. But, in spite of this, everyone who aspires to study the subject must obtain this book. E. G. R.

Some Problems in the Theory of the Luminescence of Crystals, by E. I. ADIROWITSCH. Pp. xii + 298. (Berlin : Akamedie-Verlag, 1953.) (Translated into German from the Russian.)

Luminescence studies provide a rich field of speculation for the theoretician but one in which it is dangerously easy to select data for the furtherance of a theory which later proves to be erroneous. The author has selected topics such as phosphorescence and non-radiative transitions in phosphors for detailed theoretical consideration but there is no evidence that he has avoided the pitfall above. In fact his perspective is apparently distorted by two factors, a lack of appreciation of developments, both experimental and theoretical, in the Western world and a desire (whether voluntary or not cannot be judged) to vindicate the prior and superior efforts of his fellow countrymen in this field.

In chapter I a fairly successful attempt is made to provide a rigorous definition of luminescence, distinguishing it from other emission processes such as the Raman effect and thermal radiation. This is followed by some discussion of candoluminescence and the elementary laws governing the efficiency and spectral characteristics of luminescence. Chapter II gives a historical survey of theories of decay of luminescence with an expected emphasis on the efforts of Russian workers to lay the basis of modern theories. Figure 15 of this chapter emphasizes the lack of understanding which the author has for the implications of practical data on phosphors. Many more examples of this are to be found in later chapters. His cited proofs for the ' bimolecular ' law of phosphorescence are very much dated and are not now generally accepted.

After chapter III, which gives a detailed account of the energy band model for crystals, the author returns to his main concern. This is the setting up of the ' elementary ' decay laws of phosphorescence. A multitude of experimental decay curves with theoretical ' fits ' are provided, impressive to any but those ' skilled in the art ' who realize that the decay curve reveals relatively little about fundamental processes. At the end of the chapter the studies made in the reviewer's laboratory receive castigation because they offend the Russian ' bimolecularists '. What the author never seems to realize is that the recombination of excited electrons with emission centres in phosphors is accepted by most workers in the field as involved in the luminescence decay of photoconducting sulphide phosphors. However, it is hardly ever the rate determining step in the decay, this being given by trapping processes either within centres or in the matrix lattice.

Later chapters contain more criticism of non-Soviet efforts particularly in the matter of non-radiative processes which occupy most of the later pages. However, the author's theoretical efforts in chapter VII (pages 262–3) show how erroneous his own reasoning is! He derives a temperature dependence of emission from a model which on correct solution can only give a temperature-independent emission efficiency.

The text concludes with reference to a light emission produced by cooling phosphors in the dark and an attempt to give a theory for the effect. The experiment is worthy of note but needs to be examined more carefully before a theory is postulated.

The general impression given by the text is that the author has lost perspective in the field and is promoting theories with very limited applications at the expense of ignoring a great deal of experimental data. It is surprising to find so little reference to the studies of the Eindhoven group on the nature of emission centres in sulphide phosphors, to the work in the U.S.A. and elsewhere on configurational coordinate diagrams for centres, and to the excellent experimental and theoretical work on the decay characteristics of luminescence and photoconduction in single sulphide crystals by Broser and Warminsky in Berlin. It is also surprising that current problems such as interpretation of thermoluminescence data, the effects of infra-red radiation on phosphors and the sensitization of luminescence receive little or no attention.

References to non-Russian work are all pre-1950. No subject index is provided. G. F. J. GARLICK.

Radio Astronomy, by A. C. B. LOVELL and Colleagues. Pp. 50. (London: Royal Astronomical Society, Occasional Notes, No. 16; Vol. 3, 1954.) 6*s*.

In a subject which is expanding as rapidly as Radio Astronomy, it is difficult for those outside the field to keep abreast of the developments, and any general survey is to be welcomed. In this issue of *Occasional Notes* Professor Lovell and his colleagues have combined to give ten short summarizing papers on different aspects of contemporary work. It is inevitable that a treatment of this type should lack a certain continuity, and this is particularly evident in the different standards of knowledge which the authors have assumed. Nevertheless they have succeeded in giving, to the reader without special knowledge of either astronomy or radio techniques, a stimulating account of the major advances of recent years.

The basic principles of Radio Astronomy are described in the first two chapters, where the problems of the reception of the natural emission from celestial objects, and their observation by radar methods, are discussed. In the former discussion the expression relating the noise power in an aerial situated in a radiation field is incorrect, and some of the treatment is over-simplified, but it succeeds in conveying a general picture of the experimental difficulties.

The following two contributions represent excellent accounts of the radar observations of meteors, the aurora and the Moon; both the principles underlying the observations, and the exciting results which have been obtained, particularly in the delineation of meteor radiants and velocities, are well presented.

The section describing the radio emission from the sun might well have been extended, and is in parts confusing—particularly when dealing with the emission from the undisturbed solar disc.

The remaining chapters provide a very good account of the investigation of the galactic and extra-galactic sources of radio waves, the line emission from interstellar hydrogen and the scintillation of radio stars caused by irregularities in the terrestrial ionosphere. M. RYLE.

CONTENTS OF SECTION A

PAGE

Dr. E. W. LEE. The [110] Magnetostriction of Some Single Crystals of Iron and Silicon–Iron 65

Dr. J. JACOBS. The Effect of 3pπ Electrons : Energy Levels of Ethylene . . 72

Dr. L. F. H. BOVEY. An Extension of the Analysis of the Yttrium Spectrum . . 79

Dr. J. A. POPLE. The Electronic Spectra of Aromatic Molecules : II—A Theoretical Treatment of Excited States of Alternant Hydrocarbon Molecules based on Self-Consistent Molecular Orbitals 81

Prof. D. R. BATES and Dr. G. W. GRIFFING. Inelastic Collisions between Heavy Particles : IV—Contribution of Double Transitions to certain Cross Sections including that associated with the Ionization of Hydrogen Atoms in Fast Encounters with other Hydrogen Atoms 90

Dr. G. W. GREENLEES. Excitation Functions for the Scattering of Protons and Deuterons by Be and Mg 97

Mr. N. J. PATTENDEN. The Slow Neutron Cross Section of Scandium . . 104

Mr. B. F. BAYMAN and Dr. M. A. S. ROSS. A Nuclear Emulsion Study of the Radioactive Decay of Actinium X(^{223}Ra) 110

Mr. R. WOLFE. The Theory of the Reflectivity of Metals 121

Reviews of Books 128

Contents of Section B 132

Figure 1. An electron micrograph of thin (111) lamellae in a silicon crystal. × 52 000.

Figure 2. (111) face of a silicon crystal etched in KOH for two minutes. A number of (111) lamellae terminate in etch pits. × 600.

On the Measurement of Minority Carrier Lifetime in n-Type Silicon

By J. B. ARTHUR, W. BARDSLEY, A. F. GIBSON and C. A. HOGARTH
Radar Research Establishment, Great Malvern, Worcs.

MS. received 13*th September* 1954 *and in final form* 17*th November* 1954

Abstract. A number of methods has been devised in the past few years for measuring the lifetime of minority carriers in semiconductors. With germanium all these methods give results consistent with each other. When another semiconductor, namely silicon, is studied quite large differences in the results from different methods are found. These discrepancies are described and their origin discussed. It is concluded that an uncritical application of well-established germanium technology to any other semiconductor may lead to erroneous conclusions.

§ 1. Introduction

One of the important characteristics of any sample of a semiconductor is the lifetime of minority carriers injected into the material. It is common practice in all laboratories to carry out a routine check on the carrier lifetime in all crystals produced. Apart from the fundamental information that can be obtained from lifetime measurements, the design of devices (e.g. transistors) is impossible without some information on carrier lifetime.

Various methods have been devised in the past for the measurement of carrier lifetime and employed with considerable success on crystals of germanium. The more important of these methods will now be described.

(i) *The Drift Mobility Method*, originally due to Haynes and Shockley (1951). In this method a pulse of minority carriers is injected into a filament or thin rod of the material under examination, the carriers being swept along the filament by an externally applied field and detected at a collector point some distance away. The transit time between emitter and collector may be measured directly by conventional techniques and hence the rate of decay of the carrier density determined. This is essentially a direct method, and measures the values of carrier lifetime and mobility that are operationally significant (Shockley 1951).

(ii) *Conductivity Decay Methods*. Minority carriers may be created in a filament by illumination (photoconductivity) or by injection from a rectifying contact. In either case the conductivity of the specimen is increased. The decay of the conductivity on the cessation of excitation is exponential if the original disturbance was small. Inasmuch as the decay in carrier density is observed directly, this method reproduces closely the conditions assumed in the definition of carrier lifetime. It has, therefore, been widely adopted for research purposes (e.g. Navon, Bray and Fan 1952, Moss 1953, Many 1954). As will be shown, however, care must be taken in interpreting results obtained by this method if any trapping of carriers can take place.

(iii) If the minority carrier mobility in a particular semiconductor is assumed to be known, the carrier lifetime can be obtained from the measurement of a length, rather than a time. The emitter point efficiency method (Shockley, Pearson and Haynes 1949, Hogarth 1953) and the travelling light spot methods (Valdes 1952) are representative of this group. The latter method is of great technical importance as it is widely used for routine examination of crystals. It has the great merit that the sample need not be cut to any particular shape or size. The conditions necessary to obtain valid results have been indicated by Valdes (1952).

§ 2. THE MEASUREMENT OF LIFETIME IN SILICON

When silicon crystals were first examined for transistor action in this laboratory it was quite apparent that the lifetime in all samples available was very short. The samples used were either grown here or obtained from other laboratories. In spite of the evidence of very short lifetime indicated by point-contact transistor experiments, the travelling light spot equipment indicated lifetimes of the order of 100 μsec. This equipment was known to give valid results when used with germanium crystals having lifetimes ranging from 0·2 μsec to about 200 μsec. It appeared therefore that a serious discrepancy in lifetime measurement was being obtained when silicon was used and hence a thorough examination was initiated. The study was restricted to n-type material as the results on p-type were appreciably more complex. It will be shown that all the methods described in § 1 can be made to give the same answer in n-type silicon only if certain additional precautions are taken. In view of the results obtained, together with the essential similarity between germanium and silicon, it would appear worth while to consider carefully the validity of lifetime measurements on the new semiconducting compounds which are currently receiving attention.

For the purpose of description, we shall restrict ourselves to results obtained on samples cut from one particular crystal of n-type silicon. This procedure allows for cross-correlation between the various methods used. Otherwise the results are typical of many crystals examined. In addition the crystal chosen had a lifetime sufficiently long to allow drift mobility measurements to be made and sufficiently short to show up the failures of the travelling light spot method in a convincing manner. The crystal was obtained by zone-refining high purity silicon from Johnson Matthey and Co., Ltd. The resistivity was about 30 ohm cm.

§ 3. DRIFT MOBILITY MEASUREMENTS

The drift mobility of injected carriers in both p- and n-type filaments of silicon has been measured. All the common arrangements have been used at various times, i.e. pulsed emitter, d.c. sweep field; pulsed emitter, pulsed sweep field; and d.c. emitter, pulsed sweep field. Our results are in adequate agreement with those of Haynes and Westphal (1952).

When using filaments of relatively short lifetime it has been found desirable and often essential to electroform the collector contact to increase its efficiency. For n-type silicon the collector wire contained a significant n-type impurity, following the work by Granville *et al.* (1955) on silicon point-contact transistors. The emitter contact was made from duralumin.

In order that lifetime values may be obtained from the travelling light spot experiment it is essential that the drift mobility in the sample shall be known. We include, therefore, some representative measurements on the drift mobility of holes in the crystal of n-type silicon under consideration. These results are given in figure 1 where the carrier velocity is plotted as a function of the sweeping field applied. In this example the pulsed field, pulsed emitter method was used.

In the early stages of these measurements it was observed that the output waveform was not symmetrical but skew and characteristic of trapping (Lawrance 1954). Experience with germanium at low temperatures suggested the use of background illumination, which removed the trapping and increased the mobility. The effect of the background light tends to saturate and very little increase in mobility could be obtained after that indicated by curve B in figure 1.

Familiarity with trapping in germanium at low temperatures led to the facile assumption that the trapping observed in silicon at room temperature was due to similar causes. Experiments similar to those described by Lawrance (1954) were therefore carried out and the following results obtained: (i) unlike germanium, the carrier velocity is proportional to the sweep field even under trapping conditions (cf. figure 1, curve A); (ii) unlike germanium, the carrier velocity is largely independent of emitter current under trapping conditions; (iii) unlike germanium, the carrier velocity is unaffected by magnetic fields (up to 3 kilogauss) under trapping conditions; (iv) the effects of trapping could be removed by illumination of the collector contact only, the drift space between emitter and collector remaining dark.

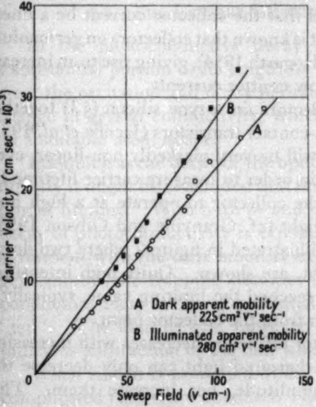


Figure 1.

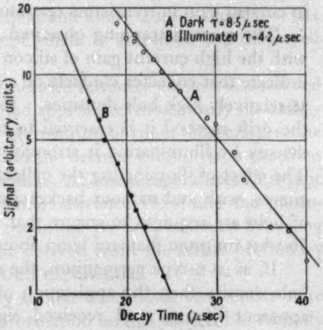


Figure 2.

From these experiments it was concluded that the trapping observed was a feature only of the collector contact and possibly due to the fact that the collector was electroformed. It was estimated that about 60% of the collected carriers were not trapped at all, even in the dark.

It follows, therefore, that the mobility of holes in n-type silicon is at least 280 $\mathrm{cm^2\,v^{-1}\,sec^{-1}}$ and probably higher. A further correction for the non-uniform resistance of the specimen should also be included (Haynes and Shockley 1951) which raises the value to about 350 cm^2 v^{-1} sec^{-1}.

§ 4. The Minority Carrier Lifetime by the Drift Method

As the holes injected at the emitter drift along the filament recombination takes place and the density of carriers arriving at the collector is proportional to $\exp(-t/\tau)$ where t is the transit time and τ the lifetime. This relation is valid only in the absence of diffusion or when an emitter pulse of sufficiently long duration is used to ensure saturation. If the emitter pulse is of very short duration the hole density is proportional to $t^{-1/2}\exp(-t/\tau)$. In practice the $t^{1/2}$ term is usually insignificant compared with the exponential, but we have included it where appropriate.

In order that the hole density at the collector, and hence the carrier lifetime, shall be measured accurately it is essential that the collector current be a linear function of the local hole concentration. It is known that collectors on germanium are not linear at small hole concentrations (Hogarth 1954), giving rise to an increase in current gain in transistors operated at low emitter currents.

The marked trapping observed at collectors on n-type silicon (§ 3) together with the high current gain of silicon point-contact transistors (Jacobs *et al.* 1953) indicate that collector contacts on silicon will be very markedly non-linear, even at relatively high hole densities. Hence, in order to measure carrier lifetime by the drift method it is essential to force the collector to operate at a high hole density by illuminating it strongly with light (cf. Granville and Gibson 1953). The effect of illuminating the collector is illustrated in figure 2 where two decay curves, with and without background light, are shown. Quite high intensities of light are required to ensure that τ has reached the limiting value, typically a 36-watt tungsten filament lamp about 5 cm from the collector point.

If, as in n-type germanium, the collector efficiency decreases with increasing hole density then the application of a background light can only decrease the apparent lifetime and received signal amplitude, not increase them. This statement is in full agreement with our experience with n-type silicon and illustrated by the results given in figure 2.

If, however, the efficiency of collectors on p-type silicon passes through a pronounced maximum at a critical hole density, as indicated by the results of Jacobs *et al.* (1953), the received signal in a drift experiment may increase and subsequently decrease when a background light is brought up. This effect has been observed in p-type material and it can be most marked. Thus, under suitable conditions, an increase in signal of 10 to 100 times can be brought about by bringing a lighted cigarette to within a few centimetres of the collector contact. Increased illumination reduces the signal to its limiting value.

To conclude, therefore, it is believed that the carrier lifetime can be determined unambiguously by the drift experiment, providing the collector is illuminated. Curve B of figure 2 is typical of the silicon crystal under consideration, the average value of ten measurements on three filaments cut from the same crystal being 3·9 μsec (extreme values obtained 3·2 μsec and 5·0 μsec). No correction has been applied for surface recombination. If the surface recombination velocity is assumed to be infinite the above figure is raised to about 4·8 μsec.

§ 5. THE CARRIER LIFETIME BY THE EMITTER POINT EFFICIENCY METHOD

As already noticed, the lifetime of injected carriers can be derived from data obtained to measure the efficiency of emitter contacts (Shockley *et al.* 1949). Inasmuch as injected carriers drift down a filament, the method is essentially similar to the drift mobility method, though appreciably less accurate. For the present purpose, however, this approach to lifetime measurement has the notable merit that no collector contact is used (only ohmic voltage probes) and hence any confusion due to non-linear hole detection is eliminated.

A filament cut from the n-type silicon crystal under discussion was measured for emitter point efficiency and value of the lifetime derived. The result was in adequate agreement with the drift method and, probably more important, the lifetime was independent of background illumination. The light did, however, substantially reduce the injection efficiency of the emitter (cf. Hogarth 1953).

§ 6. THE CARRIER LIFETIME BY PHOTOCONDUCTIVE DECAY

In the photoconductive decay method the filament is illuminated over a substantial portion of its length by short pulses of white light. In our equipment the excitation and decay periods are equal and about 2 milliseconds. The change in crystal conductivity, which is kept small, is amplified and displayed on a suitable oscilloscope.

Inasmuch as the decay of the optically generated minority carriers is observed directly it might be concluded that this method approaches most nearly the definition of lifetime. With both p- and n-type germanium at room temperature, the decay of photoconductivity is exponential and the derived lifetime is in excellent agreement with the drift mobility experiment.† At low temperatures in n-type germanium hole trapping becomes important (Lawrance 1954) and the photoconductive decay is only exponential over the initial portion, followed by a long 'tail'. The lifetime can, however, still be derived from the initial portion by judicious subtraction of the tail, which is itself exponential.

When silicon is examined under exactly similar conditions, very much more complicated results are obtained. Some information on this subject has recently been published by Haynes and Hornbeck (1954) and similar, though less exhaustive, results have been obtained here. We shall restrict ourselves to the short period decay in n-type silicon, which is all that is relevant to the present discussion.

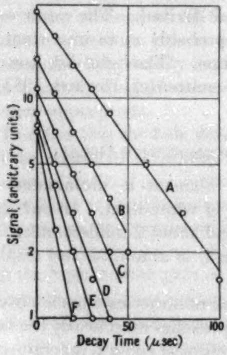
The photoconductive decay in n-type silicon is usually exponential over the first portion and this may, or may not, be followed by a tail of longer time constant. The occurrence or absence of a tail appears to be a feature of the sample preparation (e.g. the etchant used) rather than a property of the original crystal. The simplest interpretation of this result is to suppose that, of the initial minority carriers produced by the light, some are trapped and some untrapped. The initial decay then corresponds to the lifetime of free carriers, the tail corresponding to the decay of trapped carriers.

---

† This statement is not valid for intrinsic or near intrinsic germanium. As the material approaches the intrinsic condition the drift mobility and diffusion constant change markedly (cf. Arthur *et al.* 1955) and in addition the photoconductive decay ceases to be accurately exponential (Hogarth, unpublished data).

It is believed, however, that the initial portion also is affected by shallow traps, and the following evidence is submitted:

(i) The decay rate, even in samples showing no ‘ tail ’ to the decay curve, does not yield a lifetime which agrees with that from drift mobility experiments on the same specimen, without any intervening treatment. It is found that the lifetime by photoconductive decay is about ten times that from the drift method. This is a large and serious discrepancy. A typical decay curve is shown in figure 3, curve A.



Curve	Distance to background lamp (cm)	Apparent decay time (μsec)
A	∞	37
B	20	27
C	15	22·5
D	10	18·5
E	8	15
F	6	10

Figure 3.

(ii) The decay rate, and hence the apparent lifetime, is markedly affected by background illumination. This is illustrated in figure 3, curves B, C, D, E and F. This effect cannot be associated with the apparent change in lifetime with background illumination in the drift experiment as this is solely a feature of the collector contact (§§ 3, 4, 5). It can, however, be ascribed to the suppression of traps (Lawrance 1954) provided it is supposed that even the initial part of the decay is influenced by trapping. If this interpretation is correct it should be possible to reduce the apparent lifetime until, at sufficiently high intensities, a limiting value equal to the ‘ true ’ (i.e. drift) lifetime is reached. The results given in figure 3 indicate this type of behaviour.

Thus in the photoconductivity decay experiment trapping is believed to play an important part. Trapping, other than at the collector contact, is unimportant in the drift experiment because (*a*) the average time in traps, as indicated by figure 3, curve A, is 37 μsec. As the duration of the collector pulse is typically about 4 μsec, carriers arriving at the collector 37 μsec later do not materially influence the shape and magnitude of the original pulse; (*b*) since a finite time must elapse before a carrier is likely to become trapped, the proportion of carriers trapped in the photoconductivity decay experiment (2 milliseconds excitation period) must exceed the proportion trapped in a drift time (typically 20 μsec).

To conclude, therefore, it can be stated that the decay of photoconductivity in silicon does not represent a reliable method of determining the lifetime and this statement may be applicable to other semiconductors.

§ 7. THE CARRIER LIFETIME BY THE TRAVELLING LIGHT SPOT METHOD

It was quite apparent in the very early stages of work on silicon that the travelling light spot method gave values of the diffusion length (and hence lifetime) much in excess of those derived from, for example, experience in handling silicon point-contact transistors. In addition, the sensitivity distribution around the collector point was rarely exponential, though it could be analysed into two or more exponentials depending on the interpretation of the data by any particular experimenter. More confusing was the fact that, on some occasions, the signal actually increased as the light spot was moved further away from the collector, implying negative diffusion lengths!

It was quite apparent that some major effects were masking the correct sensitivity distribution and the equipment was carefully checked, using germanium crystals, for scattered light and other obvious defects. As far as could be ascertained, the equipment was not at fault. It was found, however, that the following two effects did considerably influence the diffusion length obtained:

(i) *Non-linear collection.* This is precisely the effect already discussed in § 5 and it can be removed by the use of a background light of sufficient intensity.

(ii) *Back scattered light.* With a slice of germanium about 0·5 mm thick, having a diffusion length much less than this thickness (about 0·02 mm), the signal with the collector and light source on opposite sides is about 30 times less than the signal with collector and source on the same side. The signal on the reverse side cannot be due to carrier diffusion and must be due to light of wavelength near to the absorption edge (1·8 microns) which is partially transmitted.

In the normal arrangement with light spot and collector on the same surface it follows that, if any penetrating light is scattered off the back surface of the specimen and reaches the front surface, the signal produced by it will be at least 10^3 times less than the primary signal, and therefore negligible.

Under identical conditions the back scattered signal on silicon crystals is not negligible. It is found that the signal on the reverse side of a 0·5 mm thick sample is only about three times less than the signal with collector and source on the same side. Any backscattered light will, therefore, be important when the collector is more than one or two diffusion lengths from the light spot.

§ 8. MEASUREMENTS BY THE TRAVELLING LIGHT SPOT METHOD ON n-TYPE SILICON

In order to demonstrate in a convincing manner the importance of the two effects described in the previous section, the following experiment was undertaken. A slice, about 0·5 mm thick, of the n-type silicon crystal under present consideration was mounted and the sensitivity distribution about a collector point was measured. This was repeated 45 times until the entire surface area had been sampled. The resultant sensitivity distributions were then plotted out and some selected examples are given in figure 4. The unconvincing nature of these curves is quite evident.

The entire experiment was then repeated with a powerful background light to ensure linearity of the collector and a short wavelength transmission filter over the signal source to cut out the penetrating radiation. In spite of the resultant decrease in signal strength and hence reduced accuracy, much more convincing distributions were obtained, as illustrated in figure 5.

In order to analyse the results more fully, values of apparent diffusion length L_P were obtained from every straight portion of every curve. The reciprocals of these values were then plotted as a statistical distribution (figure 6).

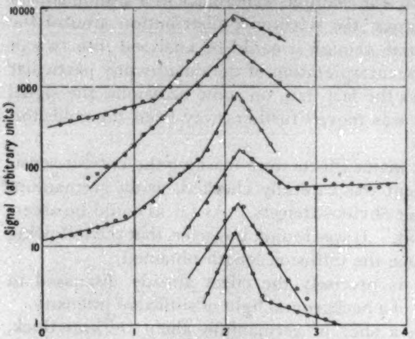


Figure 4.

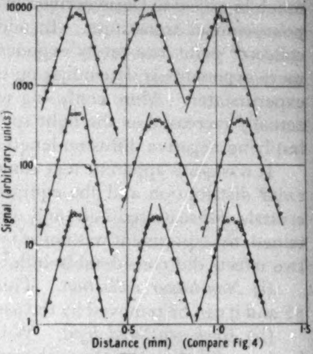


Figure 5.

The most probable value for the reciprocal of the diffusion length is about 1·5 mm^{-1} corresponding to a lifetime of about 800 μsec, the mobility being known by experiment (§ 3). As this value exceeds the drift value by 200 times, the discrepancy is very serious.

The situation is very markedly improved when the background light is applied and the signal light suitably filtered. The new distribution is given in figure 7. The following features will be noted: (i) no negative or near infinite diffusion lengths occur; (ii) the most probable value corresponds closely to the value obtained by the drift method, the latter being marked on the diagram; (iii) the spread of diffusion length values, represented by the half width divided by the value at the peak, is markedly less than before. The remaining spread (about ± 25%) may be in the measurement technique but probably represents the non-uniformity of lifetime across the specimen.

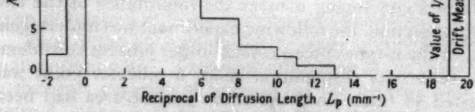


Figure 6.

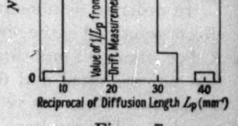


Figure 7.

To conclude, therefore, it may be stated that the travelling light spot method is valid for silicon only if certain specified precautions are taken, without which completely erroneous results may be obtained.

§ 9. CONCLUSION

Four methods of measuring the lifetime of minority carriers in n-type silicon have been studied and it has been found possible to make at least three of them agree with each other under suitable conditions. Probably the most important achievement is the elucidation of the conditions under which the travelling light spot method is valid. This method is widely used and may be used incorrectly. Our final conclusion is that care must be taken before any one method of lifetime measurement is accepted as valid when any semiconductor other than germanium is under examination.

ACKNOWLEDGMENTS

The authors are indebted to their colleagues at this Establishment for their help and advice. In particular we should like to thank Miss J. M. Shaw for her painstaking work with the travelling light spot equipment and Mr. J. B. Gunn for pointing out the possibility that the back scattered light might be important.

We are indebted to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

ARTHUR, J. B., BARDSLEY, W., BROWN, M. A. C. S., and GIBSON, A. F., 1955, *Proc. Phys. Soc.* B, **68,** 43.
GRANVILLE, J. W., BARDSLEY, W., and GIBSON, A. F., 1955, to be published.
GRANVILLE, J. W., and GIBSON, A. F., 1953, *Proc. Phys. Soc.* B, **66,** 1118.
HAYNES, J. R., and HORNBECK, J. A., 1954, *Phys. Rev.*, **94,** 1437.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81,** 835.
HAYNES, J. R., and WESTPHAL, W. C., 1952, *Phys. Rev.*, **85,** 680.
HOGARTH, C. A., 1953, *Proc. Phys. Soc.* B, **66,** 845; 1954, *Ibid.*, **67,** 636.
JACOBS, H., MATTHEI, W., and BRAND, F. A., 1953, *J. Appl. Phys.*, **24,** 1410.
LAWRANCE, R., 1954, *Proc. Phys. Soc.* B, **67,** 18.
MANY, A., 1954, *Proc. Phys. Soc.* B, **67,** 9.
MOSS, T. S., 1953, *Proc. Phys. Soc.* B, **66,** 993.
MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford: University Press), p. 131.
NAVON, D., BRAY, R., and FAN, H. Y., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 1342.
SHOCKLEY, W., 1951, *Semi-Conducting Materials*, Ed. H. K. Henisch (London: Butterworths Scientific Publications), p. 26.
SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R., 1949, *Bell Syst. Tech. J.*, **28,** 344.
VALDES, B., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 1420.

Constriction of the Positive Column of a Glow Discharge at High Pressure

By R. G. FOWLER†
University of Oklahoma

MS. received 6*th August* 1954 *and in amended form* 11*th November* 1954

Abstract. It is proposed that the constriction of the positive column into a wall-independent filament which sets in with increasing density comes about as a consequence of concentrational effects accompanying electron energy loss. As a first approximation this energy loss has been treated as a damping field acting on the electron motions only. In helium, using either M–B or Smit distributions, the constriction is predicted to lie in a range of densities in which it is indeed known to occur. A semi-empirical calculation of the critical values is made for other gases as well, with reasonable agreement with the limited data available.

§ 1. Introduction

At gas densities of less than 10^{16} molecules per cm^3, the positive column of a glow discharge in a gas free of electro-negative components normally fills the containing tube completely. The radial distribution of ions in it is known to approximate reasonably well to a simple relation first deduced by Schottky (1924). As the density of the neutral gas is increased at constant current, the column recedes from the walls and eventually becomes a filament of conducting gas. This effect is not predicted by Schottky's theory.

Attempts to establish a cause for the constriction have taken several forms. One is based on the intrinsic non-linearity of the basic differential equations, from which Schottky's theory is derived, and supposes that the phenomenon is a second-order consequence which cannot be foreseen because of the complexity of the equations. Explanations in terms of wall sheaths belong to this category. The mechanism which forms the wall sheath, however, is such that increasing density leads to a reduction rather than an increase in the thickness of the sheath, and in general it can be shown that the second approximation to the solution of these basic equations gives little more of a tendency toward constriction than the first approximation (Allis and Rose 1954).

In a second group of explanations complexities are introduced in the ion production process. Two stage ionization and volume electronic recombination both fall into this class. A production or loss term which is quadratic in the ion concentration is added to the usual linear term for single stage ionization. Such terms, while altering the distribution contours slightly in the desired direction, do not have an effect of the magnitude needed to describe the constriction (Spenke 1950). In general they require solutions in Weierstrassian elliptic functions rather than the usual circular (or Bessel) functions. It seems probable that even when the positive column finally contracts into a wall-independent filament it is unnecessary to take account of recombination since in the first place the radial ambipolar current tends to zero as density increases and in the second place the rate of ion production simultaneously falls to the very small value needed to maintain the gas conductivity in the presence of whatever trivial ion losses volume recombination inflicts.

---

† John Simon Guggenheim Memorial Fellow at the Clarendon Laboratory, Oxford, 1953–4.

In a third approach it is supposed that an increase in gas temperature at the centre of the column causes the constriction (Elenbaas 1951). It is very difficult to assess the true merits of this hypothesis because the equations do indeed become too intricate to achieve definite results. However, observation suggests that the constriction sets in in a density range well below that at which temperature gradients are of significant magnitude. Examination of the general equations shows that they take an especially simple form in two mathematical cases: first, when gas and electron temperatures are independent of position while ion concentrations vary, as might be expected at low gas densities, second, when ion concentrations are identically zero and gas temperature varies. The latter may well correspond to the outer regions of the column at high gas densities, with the ionized column surrounded by a mantle of hot neutral gas which transports heat outward to a region where convection can remove it as visualized by the theories of Elenbaas.

A fourth possible basis of explanation lies in the self-magnetic field of the column. Here calculations and observations (Thonemann and Cowhig 1951) show that the effect is indeed to be expected at very large currents (order 100 A) and low pressures (10^{-3} mm Hg), but it is well known that constricted columns can be maintained with very small currents.

A new suggestion is advanced in this paper. It is proposed that inelastic collisions have, in addition to the generally recognized energy losses which they impose on the electrons in the column, a concentrating action (or, more explicitly, an anti-dispersing action) on the electrons because of the decelerations occurring during the collisions.

§ 2. COLLISION DAMPING OF THE ELECTRON MOTIONS

In a gas in thermal equilibrium, all collision processes whether elastic or inelastic are balanced in detail. The ionized gas in the positive column of a glow, however, is never in thermal equilibrium but in a steady state in which the energy received from the static electromagnetic field is balanced by the radiation loss and the heat transferred to the surroundings.

In this steady state, the electrons have a ' temperature ' which is much higher than that of the gas with which they are interacting. Thus, on an average they deliver more energy to the gas molecules even during elastic collisions than they receive in return in balancing collisions. This energy loss is more pronounced the greater the disparity between the electron and gas temperature. For inelastic collisions, moreover, the energy losses will be uncompensated by reverse collisions under all practical conditions. To provide conditions in which they would be balanced by collisions alone is nearly if not wholly impossible in terrestrial discharges.

Losses of energy of the two types mentioned result in a deceleration of the electron which, as far as an electron gas drifting or diffusing in a field-free space is concerned, is similar in macroscopic effects to that which would be experienced in a retarding force field. Dimensionally, and in order of magnitude, the force field will be of the form

$$F \propto \sum_{m} eV_m \sigma_m N \qquad \ldots\ldots(1)$$

where eV_m is the mean energy loss in the collision of type m; σ_m is the mean cross section of the collision; N is the number of neutral particles; and the summation proceeds over all types of collisions in which there is a statistically unbalanced energy loss. Eventually it should be possible to evaluate the proportionality constant for this relation properly from the collision integral. For our purposes the constant will be taken as unity, and a damping field defined as $E_k = F/e$.

§ 3. APPLICATION OF A DAMPING FIELD TO THE THEORY OF POSITIVE COLUMN

The first-order equations of transport theory which govern diffusion of positive and negative ions in a force field in the absence of recombination and of temperature gradients, and assuming production of ions by single collisions, are

$$-\operatorname{div} j_+ + \nu n_- = 0 \qquad \ldots\ldots(2)$$

$$-\operatorname{div} j_- + \nu n_- = 0 \qquad \ldots\ldots(3)$$

$$j_+ = -D_+(\operatorname{grad} n_+ - F_+ n_+/\boldsymbol{k}T_+) \qquad \ldots\ldots(4)$$

$$j_- = -D_-(\operatorname{grad} n_- - F_- n_-/\boldsymbol{k}T_-). \qquad \ldots\ldots(5)$$

Here (with appropriate subscripts) j represents particle current density, ν the rate of ionization, D the diffusion coefficient, n ion concentration, T the temperature, and F a force field. To these is added the Poisson equation

$$\operatorname{div} E_s = 4\pi e(n_+ - n_-). \qquad \ldots\ldots(6)$$

Introduction of the damping field (which we have called E_k in distinction from the coulomb field E_s) results in an interpretation of the forces in (4) and (5) as

$$F_+ = eE_s \qquad \ldots\ldots(7)$$

and

$$F_- = -eE_s - eE_k. \qquad \ldots\ldots(8)$$

It is an essential point that the Newtonian reaction from the force eE_k is not upon the positive ions, but upon the neutral.

Employing the principle of local neutrality which leads to the usual ambipolar approximation, viz. $n_+(x, y, z) = n_-(x, y, z) = n$, (2) and (3) become

$$D_-(\nabla^2 n + \operatorname{div} eE_s n/\boldsymbol{k}T_- + \operatorname{div} eE_k n/\boldsymbol{k}T) + \nu n = 0 \qquad \ldots\ldots(9)$$

$$D_+(\nabla^2 n - \operatorname{div} eE_s n/\boldsymbol{k}T) + \nu n = 0 \qquad \ldots\ldots(10)$$

where T is the gas temperature, and $T_+ = T$. Subtracting (10) from (9) and integrating, we obtain the net particle current density as

$$i/e = (D_- - D_+)\operatorname{grad} n + (D_- T + D_+ T_-)eE_s n/\boldsymbol{k}TT_- + D_- eE_k n/\boldsymbol{k}T_- \qquad \ldots\ldots(11)$$

where i is the current density, and will be assumed to have only one component. Introducing this into either (9) or (11) leads to

$$\nabla^2 n + \operatorname{div} eE_k n/\boldsymbol{k}(T_- + T) + \nu n/D = 0 \qquad \ldots\ldots(12)$$

where D is the ambipolar diffusion coefficient, defined as

$$D = D_+ D_-(T + T_-)/(D_+ T_- + D_- T) \sim D_+(T + T_-)/T. \qquad \ldots\ldots(13)$$

Properly defined, E_k would be a function of the peculiar velocity of the electrons. We have already made the physical assumption, however, that it is governed by (1), and we now for mathematical simplicity make the further assumption that it is constant in magnitude.

Restricting ourselves to the case of a discharge between parallel walls $2a$ cm apart, with the current density directed parallel to the walls, the solution of (12) is a damped cosine function:

$$n = Ae^{-\beta x}\cos\{(\gamma^2-\beta^2)^{1/2}x + B\} \qquad \ldots\ldots(14)$$

where $$2\beta = eE_k/\boldsymbol{k}(T+T_-), \text{ and } \gamma^2 = \nu/D. \qquad \ldots\ldots(15)$$

In the usual first approximation the boundary conditions are $dn/dx=0$ and $n=n_0$ at $x=0$. Because the ion concentration is retricted to positive values, we have

$$(\gamma^2-\beta^2)^{1/2}a + B \leqslant \pi/2 \qquad \ldots\ldots(16)$$

where $x=a$ at the tube wall. It should be noted that the solution only applies to the half of the discharge space lying between $x=0$ and $x=a$.

The solution is subject to additional requirements on the energy density balance which fixes the component of electrostatic field parallel with the walls, and on the total discharge current which determines the value of the constant B, but the occurrence of the constriction depends upon the quantities β and γ. Because the approach of β to γ is governed by the electron temperature T_-, and hence only indirectly by the gas density, and since T_- changes quite slowly at large densities, the distribution may be expected to preserve its limiting form after the constriction sets in. While other processes such as heat conduction will become important after the constriction has increased to any extent, in the present hypothesis, the constriction occurs because the electrons tend to pile up in the centre of the tube as a result of their energy losses. The criterion for the constriction can be interpreted physically as being that value of electron temperature at which the average distance within which the electron cloud would lose its thermal energy (at the rate of loss appropriate to the given temperature) is equal to the so-called diffusion length. This condition is characteristic of the gas under investigation.

Computations of the quantities β and γ have been made in helium, using Maxwell–Boltzmann statistics, and are given in figure 1. The various contributions to E_k are given in figure 2.

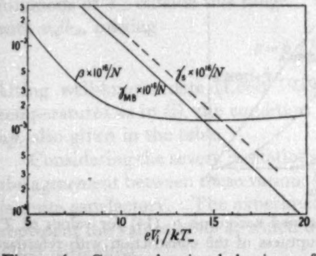


Figure 1. Curves showing behaviour of the parameters β and γ as functions of electron temperature T_-. γ_{MB} is the result of using a Maxwell–Boltzmann distribution, while γ_{S} is computed and extrapolated from Dunlop's calculations of Townsend's ionization coefficient using the Smit distribution. eV_i is the ionization energy of the gas.

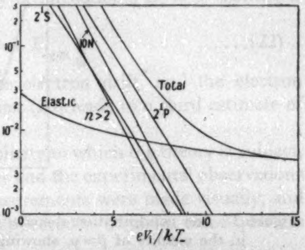


Figure 2. Various contributions to the damping field E_k in helium arising from ionization loss, elastic loss, excitation loss to 2^1S, and to 2^1P, and to all other singlet excitations lumped together. Losses to the triplet state 2^3S will become very important at small values of $eV_i/\boldsymbol{k}T_-$. E_k is given in units of cm^3 volt cm^{-1}.

Theoretical values of the cross sections have been used in view of the uncertainty about most of the available experimental data. For the elastic cross section in helium the expression

$$\sigma = 7{\cdot}8 \times 10^{-15}/(V + 11{\cdot}2) \qquad \ldots\ldots(17)$$

was used as a reasonable approximation to the curve given by Massey and Burhop (1952). For ionization and excitation the form

$$\sigma = (\kappa/V) \log V/V_c \qquad \ldots\ldots(18)$$

was chosen, although it is known to be in some disagreement with experiment at low energies. V is the electron energy in electron volts; V_c is the proper critical potential; and κ was chosen from the data given by Mott and Massey (1933) for helium.

It is found that constriction should occur in helium at an electron temperature of 20 000°K, which corresponds to a value of $ap \simeq 30$ cm × mm Hg using von Engel and Steenbeck's theory (1934) of the electron temperature. Some alteration of this value is to be expected if a Smit distribution had been used to calculate both β and γ, although the former is certainly more sensitive than the latter. Using the data for the ionization coefficient in helium obtained by Dunlop (1949, see also Abdelnabi and Massey 1953) with the use of the Smit distribution, in an effort to improve the value of γ alone, one obtains $ap = 40$.

Since both β and γ are proportional to the neutral concentration, the constriction obeys the similarity principle, and so constriction of a column will occur at lower densities in tubes of larger diameter. This prediction is fully borne out by experiments in air. That the constriction depends strongly on the electron temperature is shown by the curves of figure 3 in which the distribution of ion concentration across the discharge is given in helium for values of β and γ approaching the critical value.

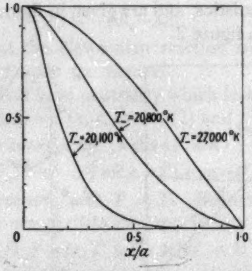


Figure 3. Ion concentration distributions computed according to (14) for values of T_- in the vicinity of $\beta = \gamma$, showing the abruptness of the constriction with reference to changes of electron temperature.

Three empirical approaches can be made in the problem of estimating the constriction condition, each based upon an aspect of E_k. (i) The longitudinal field in the positive column is governed in magnitude by the energy balance between wall losses in electron energy and volume losses, which latter we have proposed can be treated as the damping field E_k. As the constriction sets in, the wall losses will approach zero, so that measurements of the longitudinal field will tend toward E_k. Since the constriction condition is

$$eE_k/2\boldsymbol{k}T_- = (\nu/D)^{1/2} \qquad \ldots\ldots(19)$$

we may make use of the temperatures calculated by von Engel and Steenbeck on the assumption that $(\nu/D)^{1/2} = 2{\cdot}405/a$ to find a value of *ap* which satisfies this equation. These results are given for several values in the table. The method implies that the ionization is by direct electron impact, and hence is open to

Constriction Values

Gas	He	A	Ne	Air	H_2
ap (M–B theory)	30	—	—	—	—
ap (Smit theory)	40	—	—	—	—
ap (longitudinal field)	8	10	31	—	0·5
ap (electron temperature)	6	27	34	—	1
ap (electron drift)	18	6	39	6	4
ap (experimental)	45	35	25	8	3

large uncertainties. (ii) Electron temperature has been measured directly as a function of field in the experiments of Townsend (1925). Again, at the constriction onset the damping field will be equal to the entire applied field, and so to the approximation that

$$eE_k/2\boldsymbol{k}T_- = 2{\cdot}405/a, \qquad \ldots\ldots(20)$$

a second value of *ap* can be found. (iii) The drift of electrons at moderate speeds (within which range the constrictions lie) can be expressed by the empirical mobility relation

$$v = v_0 + b_- E_s \qquad \ldots\ldots(21)$$

where v_0 and b_- are constants over a wide range of E_s, and may be thought of as functions of T_- outside this range. One way of introducing E_k is to identify it with v_0/b_-, making

$$v = b_-(E_k + E_s) = b_- E. \qquad \ldots\ldots(22)$$

Using well-known data (Loeb 1939) on the electron drift, and the electron temperatures as in (i), the condition in relation (20) leads to a third estimate of *ap*, also given in the table.

Considering the severe limitations of simplicity to which the theory is subject, the agreement between these various estimates and the experimental observations is quite satisfactory. The experimental measurements were made visually, and represent the values of *ap* at which the tube appeared to be about three-quarters filled with glow.

While the entire theory can be looked upon as in (iii) as a mere empirical correction of the mobility equation, it is believed that the generalized treatment as a damping field is a superior viewpoint which will clarify other situations as well. The suggestion embodied in this paper is that although the electron energy losses have heretofore been duly considered in their effect on electron temperature and longitudinal field, no allowance has been made for their effect on particle distribution.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. A. von Engel for calling his attention to this problem and for continued interest and criticism. He also wishes to thank Professor Lord Cherwell for extending to him the facilities of the Clarendon Laboratory. The work was carried out as part of a fellowship programme under the John Simon Guggenheim Foundation, and with the added support of the Regents of the University of Oklahoma, to both of which the author tenders grateful acknowledgment.

REFERENCES

ABDELNABI, I., and MASSEY, H. S. W., 1953, *Proc. Phys. Soc.* A, **66,** 288.
ALLIS, W. P., and ROSE, D. J., 1954, *Phys. Rev.*, **93,** 84.
DUNLOP, S. H., 1949, *Nature, Lond.*, **164,** 452.
ELENBAAS, W., 1951, *The High Pressure Mercury Vapour Discharge* (Amsterdam : North Holland Publishing Co.).
VON ENGEL, A., and STEENBECK, M., 1934, *Elektrische Gasentladungen* (Berlin : Springer).
LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York : John Wiley).
MASSEY, H. S. W., and BURHOP, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford : University Press).
MOTT, N. F., and MASSEY, H. S. W., 1933, *Theory of Atomic Collisions* (Oxford : University Press).
SCHOTTKY, W., 1924, *Phys. Z.*, **25,** 342, 635.
SPENKE, E., 1950, *Z. Phys.*, **127,** 221.
THONEMANN, P. C., and COWHIG, W. T., 1951, *Proc. Phys. Soc.* B, **64,** 345.
TOWNSEND, J. S., 1925, *J. Franklin Inst.*, **200,** 563.

The Growth of Electrodeless Discharges in Hydrogen

BY G. FRANCIS
Clarendon Laboratory, Oxford

*Communicated by A. von Engel; MS. received 13th August 1954
and in amended form 18th October 1954*

Abstract. Measurements have been made on the starting potential, current, and rate of growth of current of electrodeless discharges in hydrogen. The gas was contained in cylindrical glass or quartz vessels placed between plane parallel electrodes. The wavelength was varied from 50 m to 6×10^6 m (frequency range 6 Mc/s to 50 c/s), and the gas pressure from 1 to 76 mm Hg.

At short wavelengths the starting potential and current are low, and the discharge develops slowly. At a critical 'cut off' wavelength the starting potential rises abruptly and then remains almost constant to the longest wavelengths. At wavelengths just greater than cut-off the current is large, grows rapidly and flows continuously: at all longer wavelengths a smaller current flows in a brief pulse which occurs near each peak of the applied field, and is of constant height and shape.

The growth of the pulses is thought to be due to ionization in the gas, and electron emission from the walls of the vessel by photons. A pulse ceases when wall and space charges sufficiently reduce the field in the gas.

Pulses occur so regularly that electrons must be left over from one pulse to enable the next to start. They are loosely bound on the 'anode' wall at the end of one pulse and easily pulled off when the field reverses. This emission occurs in quite small fields ($<1000\,\mathrm{v\,cm^{-1}}$), and is distinct from normal field emission.

This idea has been tested using alternating square wave fields of very long period, and by removing the wall charges.

---

§ 1. INTRODUCTION

THE present work was undertaken in order to explain the mechanism of high-frequency discharges in hydrogen over a wide range of applied frequencies and gas pressures. Spectroscopically pure gas is contained in glass and quartz vessels placed between plane parallel external electrodes, and both the starting field and the current are being measured.

The experiments are an extension of the technique used recently in work on the growth of electrodeless discharges at low pressures (Francis and von Engel 1953). By using a bridge method to eliminate circuit currents the rate of growth of the current actually carried by a discharge can be measured, and this gives information about the processes causing breakdown which could hardly be obtained in any other way. It is sometimes possible to imagine several different mechanisms which explain equally well why a discharge starts in a given applied field, but which lead to quite different rates of growth and final equilibrium currents, measurement of which enables us to decide which is the right explanation. By the current through the discharge we mean that additional current which flows between the external electrodes and around their associated circuit when the discharge starts, not counting the capacity current which flows all the time.

Measurements have been made in hydrogen in the pressure range 1 to 76 mm Hg, and over a wavelength range of 50 m to 6×10^6 m (frequency range 6 Mc/s to 50 c/s).

§ 2. The Apparatus

Cylindrical vessels were used, of two different lengths, 1·8 cm and 5·8 cm (internal), all having a diameter of 3 cm. Two vessels (one of each size) were of soda glass, two of Pyrex, and there was also one quartz vessel 1·8 cm long. Each vessel was placed in turn between plane parallel electrodes with its axis parallel to the electric field. High-frequency potential was picked up from an oscillator by means of a parallel resonant circuit and applied across the electrode system (figure 1). At wavelengths longer than 50 m ($f < 6$ Mc/s) symmetry of the electrodes with respect to earth was found to be no longer necessary, so that one side of the electrode system could be earthed, which was a great convenience. At wavelengths longer than 15 000 m ($f < 20$ kc/s) the potential was applied directly to the electrodes from a beat frequency oscillator via an amplifier and transformer. The peak voltage was measured by a suitable valve voltmeter.

Gas pressures above 3 mm Hg were measured with a mercury manometer, and below 3 mm Hg with an oil manometer, both isolated from the discharge vessel by a liquid air trap. The walls of the vessels were cleaned by heating them during continuous pumping, and by pumping out the gas while a high-frequency discharge was running. Readings were taken only when reproducible results were obtained over periods of days or weeks.

§ 3. The Current Measuring Circuit

3.1. *Bridge Circuit and Amplifier*

If a discharge occurs in a glass vessel containing gas placed between two external electrodes across which is applied a high-frequency potential two currents flow between the electrodes. One is the normal capacity current i_c which flows whether a discharge occurs or not, the other is the current i_g carried by the discharge. It is found that over an appreciable range of frequencies the discharge current is negligible compared with the capacity current, and in order to measure it a bridge current is necessary to eliminate this capacity current.

The circuit previously developed for similar experiments on low pressure electrodeless discharges at rather higher frequencies ($\lambda \simeq 20$ m) (Francis and von Engel 1953), using small transformers, was found unsatisfactory for this work, due to stray pick-up with the larger potentials necessary to start these discharges. The discriminating property necessary in the bridge remains, however, practically unchanged, i.e. it must be able to measure 1 part of discharge current in the presence of about 100 parts of capacity current; these, of course, are the worst conditions.

The bridge circuit shown in figure 1 was found to work best. A and B are two sets of electrodes constructed as nearly identical as possible, each in the form of two parallel circular discs surrounded by guard rings. A cylindrical glass or quartz vessel containing the gas is placed in A. The guard rings preserve a uniform electric field whilst all the discharge current, with the least possible capacity current flows across the inner discs. C compensates for any slight difference between the capacities of systems A and B. R_1 and R_2 are so small ($\simeq 100\ \Omega$) that the voltage drop across them is negligible compared with the voltage across the whole electrode system. Before a discharge starts the current i_c in each arm of the bridge is determined entirely by the capacity of the electrodes and by adjusting C and R_2 the potentials at the points X and Y can be made exactly equal both in magnitude and phase. R_2 consists of a small and a large resistance in parallel, the latter acting as a fine control. If a discharge occurs in the vessel the additional current i_g causes a potential difference $i_g R_1$ to appear across XY, and this is independent of the frequency of the applied field.

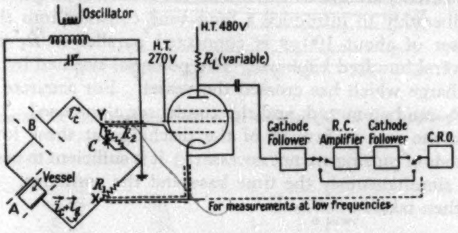


Figure 1. The bridge circuit and amplifier. Valve CV 138 or equivalent. $R_1 = 100\ \Omega$ when bridge circuit necessary, up to 5 kΩ at low frequencies.

The amplifier circuit is shown in outline in figure 1. The displacement of the oscilloscope trace is proportional to the potential difference across XY, and thence to the current flowing through the discharge. The output from the amplifier can be rectified if required to display the envelope of the peak current flowing through the discharge. By pulsing the high-frequency oscillator and synchronizing the time base, the oscilloscope is made to show the growth of the discharge current with time.

3.2. *Procedure*

The bridge is first balanced with slightly less than starting potential applied. The output of the oscillator is varied to ensure that this position of balance does not alter, and then increased very slowly until the discharge starts. The trace of the current appears on the screen and is photographed. A similar procedure using a pulsed oscillator shows the growth of the discharge current on the screen. A fraction of the voltage across the eletrodes is also displayed on the other beam of a double beam oscilloscope so that a simultaneous record of voltage and current is obtained. The time interval between successive pulses can be varied between 100 μsec and 0·1 sec or single shot photographs can be taken, and the length of a pulse can be varied from 10 μsec to 500 μsec.

3.3. *Calibration*

The displacement of the current trace is calibrated by applying known potentials from a signal generator between grid and earth of valve V_1 and measuring the movement of the trace. The signal generator is tuned to the same frequency as that being used, although over a very large range the sensitivity of the apparatus is independent of frequency.

3.4. *Circuit used for Low Frequencies*

At frequencies below 30 kc/s ($\lambda > 10\,000$ m) the capacity current across the electrodes is negligible, and the bridge circuit is unnecessary. The potential developed across R_1 is then entirely due to the discharge current, and can be displayed directly on the oscilloscope. For this purpose R_1 can be increased up to 5 kΩ without producing too great a voltage drop or too long a time constant.

At low frequencies the discharge current is found to flow in brief bursts which last a very short time compared with the period of the applied field. These sharp pulses of current are difficult to see on the oscilloscope and it is then more convenient deliberately to introduce a long time constant into the circuit. A known condenser of about 100 pF is connected parallel to R_1 which is itself increased to several hundred kilohms. The potential attained by the condenser measures the charge which has crossed the vessel. For accurate measurement of the charge R_1 can be omitted, and the condenser alone used.

To examine the rate of growth of the discharge at these low frequencies electronic methods of pulsing are not necessary; it is sufficient to use a mechanical switch to start simultaneously the time base and the oscillator. A single shot photograph is then taken.

§ 4. EXPERIMENTAL RESULTS

4.1. *Variation with Wavelength*

The results of a large number of experiments on hydrogen discharges over the wavelength range from 50 m to 6×10^6 m ($f = 6$ Mc/s to 50 c/s) and at pressures between 1 and 76 mm Hg in vessels of length 1·8 cm are given in figures 2, 3 and 4. These curves show how the starting field, final equilibrium current and the rate of rise of current vary with the wavelength.

The most noticeable feature of the curves is the sharp discontinuity in the region of wavelength about 100 m previously observed and explained by Gill and von Engel (1949), but now shown to persist up to higher pressures. Beyond cut-off, i.e. at longer wavelengths, the constancy of the starting field contrasts with the behaviour of neon (Harries and von Engel 1954) in which the starting potential increases towards longer wavelengths in several steps.

At very long wavelengths ($\lambda \geqslant 10\,000$ m approximately) the current flows in pulses which occur at or near the peak of the applied voltage. With the equipment available it was not possible to photograph these pulses accurately enough to measure the rate of rise of current but from visual observation it did not seem to vary with wavelength.

4.2. *Effect of Pressure*

At pressures above 25 mm Hg the value of X/p necessary to start these discharges in hydrogen decreases slowly with increasing pressure, and as later experiments show is equal to the value for d.c. discharges under the same conditions. The power available from the oscillators set an upper limit to the pressure which could be used. If V, the starting potential, is plotted against the product pd (pressure × length of vessel) a straight line results.

4.3. *Effect of Pulse Length and Pulse Interval*

Great care was taken in measuring the current to ensure that its rate of growth was not affected by any ionization already present in the vessel. Most of the results given here were obtained using a regular sequence of repeated pulses. The pulse length and separation were chosen in such a range that the starting potential and current did not vary with these parameters.

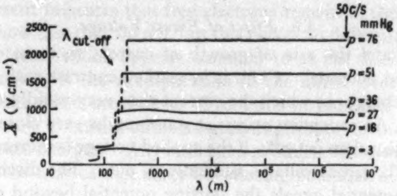


Figure 2. Variation of the starting field with wavelength of the applied field (d=1·8 cm).

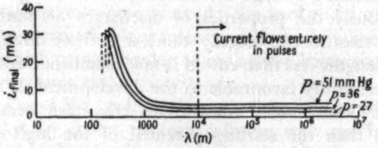


Figure 3. Variation of the field equilibrium discharge current with wavelength (d=1·8 cm).

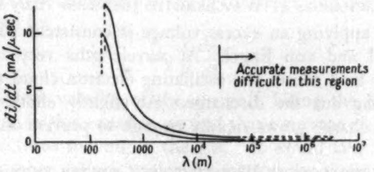


Figure 4. Variation of the rate of growth of current with wavelength (d=1·8 cm).

Sometimes in order to obtain a steady trace free from violent statistical fluctuations the vessel was irradiated with ultra-violet light from a mercury lamp. Electrons are emitted mainly from the walls of the vessel, little light being absorbed in the gas. The lamp was placed close enough to reduce the fluctuations, but not so close as to affect the rate of growth.

4.4. *Effect of the Material of the Vessel*

Vessels of different materials, Pyrex glass, soft glass and quartz were used all having identical internal dimensions. Only small differences in the starting field and the current were observed.

4.5. *Appearance of the Discharge*

The intensity of light from the discharge varies with the wavelength in roughly the same manner as the current, the glow being intensely bright and concentrated at wavelengths just longer than cut-off. At very long wavelengths it is so dim as to be hardly visible. As the wavelength is increased in the region beyond cut-off very bright layers near the end walls become steadily wider and more diffuse, although never becoming a uniform glow in the vessel.

4.6. *Effect of Excess Voltage*

The current and the rate of growth of current increase only slowly with increasing applied potential. (This is in marked contrast with the behaviour of low pressure discharges in which the current rises very rapidly with even a small excess voltage.) An exception occurs at wavelengths very close to (within a few metres) but just less than cut-off; if the applied voltage is increased the discharge will, at about 50% excess voltage, suddenly go out. The discharge starts again only when the potential equals the starting potential beyond cut-off, and then exhibits *all* the characteristics (large current, intense light) of those discharges. A similar result is observed at slightly shorter wavelengths, though still close to cut-off; with increasing voltage the discharge does not go out, but suddenly shows simultaneously the properties of discharges on both sides of cut-off. The current too rises in two stages; the low current discharge which always develops at wavelengths less than cut-off is first established and the space charges produced are presumably favourable to the development of the second type of discharge, which occurs soon after. This transition occurs at a potential considerably less than the starting potential of the high current discharge. Figure 5 (Plate) shows this effect, and some typical oscillograms of discharge at various wavelengths.

§ 5. DISCHARGES AT WAVELENGTHS JUST LESS THAN CUT-OFF

The effect of applying an excess voltage is consistent with the mechanism proposed by Gill and von Engel. At wavelengths very close to cut-off an increased applied field causes the oscillating electron cloud to be swept to the walls, thus putting out the discharge. At slightly shorter wavelengths the central ion space charge grows rapidly enough to prevent this.

§ 6. DISCHARGES AT WAVELENGTHS GREATER THAN CUT-OFF

6.1. *Deductions from Change of General Properties with Wavelength*

Figures 2, 3 and 4 show a constant starting field over the whole range, from cut-off down to the lowest frequencies, especially at the higher gas pressures, and similarly for the current over most of the range.

The starting field depends upon the most stringent condition which has to be satisfied in order that a discharge may develop, while the rate of growth of current and its final value tell us something of the development of the discharge and its equilibrium state. A discharge develops in several stages, from the multiplication of an initial chance electron to the final equilibrium state. Any one of these intermediate stages may be such as to obstruct further development of the discharge, and the starting field is that necessary just to sustain multiplication during the most difficult stage. Clearly one is not able to predict with certainty what the starting field of any given discharge will be without knowing the stages in its development.

The constancy of the starting field shows that the condition (i.e. the mechanism) for starting is probably the same for all discharges at wavelengths greater than cut-off. The discharges themselves, however, develop in different ways, the current changes steadily from being continuous and roughly sinusoidal at short wavelengths to pulses at longer wavelengths as well as decreasing in magnitude. A continuous current indicates that free electrons remain in the gas as the field reverses, pulses show that they do not, but this makes no difference to the starting field. The only factor which appears common to all the discharges, whatever the applied frequency, is the multiplication of the initial electron when space charges and all other effects are absent. From this we conclude that the starting field is that necessary to ensure cumulative multiplication of the initial electron in a gap free of space charges. Any ionization products left over in the gas from one half cycle to the next cannot affect the starting potential.

6.2. *Growth and Regularity of Current Pulses*

At most frequencies considered here the current flows only in brief pulses near the peak of the applied field, each pulse being separated by a relatively long interval. Two things have to be explained: the very rapid growth of current during a pulse, and the remarkable regularity with which pulses of opposite sign follow each other, which cannot be due to each pulse being started by a chance electron, e.g. from a cosmic ray.

The measured value of X/p ($\simeq 35$) at the peak of the field is sufficient to produce appreciable ionization by direct electron collision with gas molecules and at these pressures the multiplication is properly described by Townsend's ionization coefficient α. If we assume that the discharge starts from one electron then one avalanche would not account for the observed peak current of a few milliamperes. The relation between the charge q moving in the vessel and the current i flowing in the circuit is given by

$$i = q\frac{X}{V}\frac{dx}{dt} \qquad \ldots\ldots(1)$$

where dx/dt is the drift velocity of the charge. Using the known values for electrons in hydrogen in this equation we find that a current of 1 mA corresponds to about 2×10^9 electrons moving in the gas. It would take many avalanches to produce that number. Each avalanche takes only about 2×10^{-7} sec, this being the transit time of electrons along the length of the vessel, and so we conclude that the observed pulse of current rising to 1 mA in about a microsecond is the result of a rapid succession of avalanches. A very fast secondary process is necessary, so that some agency created in one avalanche triggers the next avalanche within about 10^{-8} second. Light is the only agency which acts so quickly and since photoionization in the gas is negligible at these pressures photoelectric emission from the ' cathode ' wall of the vessel must occur.

Multiplication of electrons cannot continue indefinitely. During the growth of a current pulse they are swept rapidly on to the ' anode ' wall of the vessel, where they stay. (It must be remembered that these are electrodeless discharges and the walls of the vessel are insulators.) Positive ions are also produced in the gas, but do not move far during the growth of the avalanches, and the reverse field thus set up eventually stops further multiplication. The current decreases rapidly, and that marks the end of the pulse. However, there remain a large electron wall charge, a large positive ion space charge, and other products, for example gas atoms in excited states, and long-lived photons. The effect of one or more of these products must constitute another secondary process which after a relatively long interval triggers the next current pulse, in the reverse direction. Also since the starting field and the current are so constant with wavelength this process must be largely independent of time over an appreciable range.

We now consider in detail the fast and the slow secondary processes.

6.3. *The Fast Secondary Mechanism and the Condition for Starting*

The process by which an initial electron is multiplied gives us the starting condition of these discharges. The applied field can be considered constant during the short transit time of an electron and so the number of ions produced from one electron starting at the cathode wall is $e^{\alpha d}-1$. Now assume that

$$z=\frac{\text{No. of photons}}{\text{No. of ions}}\text{ produced during an avalanche.} \qquad \ldots\ldots(2)$$

Most of the photons will originate from regions near the anode wall, but only a fraction f of them will hit the cathode wall, this fraction being given approximately by the solid angle which the cathode wall subtends at the centre of the anode wall. If γ_p is the photoelectric yield the starting condition is

$$\gamma_p fz(e^{\alpha d}-1)=1. \qquad \ldots\ldots(3)$$

Both α and z depend upon X/p, so that it is not a simple matter to deduce the starting field. Moreover, both z and γ_p are largely unknown quantities, although from the scant evidence in the literature it is possible to make rough calculations of z for hydrogen. However, we can take the observed value of X/p at the starting field, insert values for α and z, and see whether the value for the photoelectric yield is reasonable.

For the smaller vessels ($d=1\cdot8$ cm, diameter 3 cm) with hydrogen at a pressure of 27 mm Hg, $X/p=39\cdot1$. The corresponding α/p, from Hale's measurements, is $\alpha/p=0\cdot125$, hence $d=6\cdot1$. These values give

$$\gamma_p fz=2\times10^{-3}. \qquad \ldots\ldots(4)$$

Simple geometry gives $f=1/10$. Although most photons are produced in a layer near the end wall, which reduces f, other photons are produced along the axis of the tube and a larger fraction of these get back to the cathode wall and so $f=1/10$ seems a reasonable average. There is no point in a more accurate calculation, as we shall now show in calculating the ratio z.

We can assume that the energy distribution of electrons in hydrogen is Maxwellian, provided X/p is less than about 50 (Emeléus *et al.* 1936). Then the ratio

$$\frac{\text{No. of excitations}}{\text{No. of ionizations}}=\frac{\sum\int_{\epsilon_0}^{\infty}P_\epsilon\exp(-\epsilon/\bar{\epsilon})\,dv}{\int_{\epsilon_1}^{\infty}P_1\exp(-\epsilon/\bar{\epsilon})\,dv} \qquad \ldots\ldots(5)$$

where P_ϵ, P_1 are the probabilities of excitation and ionization, ϵ_0, ϵ_1 are the excitation and ionization energies respectively, $\bar{\epsilon}$ is the mean energy, and v the velocity.

We require only those photons of energy above a few electron volts, i.e. those able to release photoelectrons. Thus approximately z, the ratio of useful photons to ions, is equal to the above ratio. The only known value of the probability of excitation in molecular hydrogen is of the resonance level (12·8 ev) (Whiddington and Jones 1928), and Whiddington suggests an average value of about 2% for the total excitation probability. The average value of P_i is 10%, $\bar{\epsilon}$ is known as a function of X/p in hydrogen from the measurements of Brose (1925), and is equal to 4·3 ev when $X/p=39$. Hence

$$z \simeq \frac{\bar{P}_e \int_{12\cdot 8}^{\infty} \exp(-\epsilon/\bar{\epsilon})\, dv}{\bar{P}_i \int_{15\cdot 5}^{\infty} \exp(-\epsilon/\bar{\epsilon})\, dv}. \qquad \ldots\ldots(6)$$

The above values give $z \simeq 0{\cdot}5$.

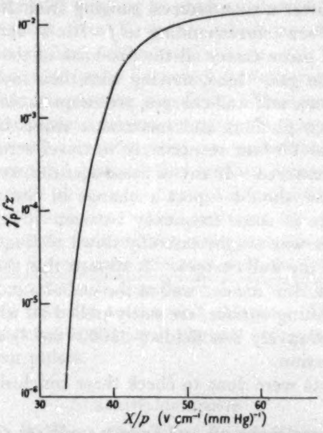


Figure 6. Variation of the fast secondary mechanism with X/p.

Substituting f and z in equation (3) we find the photoelectric yield $\gamma_p \simeq 4\times10^{-2}$. This value is quite reasonable in order of magnitude, and agrees with much evidence (Wainfan *et al.* 1953, Harries and von Engel 1954) which points to quite high values of photoelectric yields by quanta of reasonably short wavelength (see also Penning and Moubis 1935).

Several other points about these measurements are important: (i) The curve of V_s against pd is almost identical with that obtained by other workers using d.c. and metal electrodes. (ii) The values of $\gamma_p fz$ calculated at different values of X/p show a marked variation (see figure 6). The curve is similar to that obtained by Llewellyn Jones and Davies (1951), although their variation was most marked at $X/p \simeq 150$. However, they used metal electrodes. The variation is presumably due to different energy photons being produced under different exciting conditions. (iii) Equation (3) shows that the starting condition, if it is due to photons, should depend markedly on the factor f, i.e. on the geometry of the vessel. If a long vessel ($d=5{\cdot}8$ cm) is used instead of the short vessel ($d=1{\cdot}8$ cm), but is of the same diameter (3 cm), f is reduced from 1/10 to 1/60, which should necessitate an increased starting potential for the same pd. However, a quite small increase in X/p produces a large increase in z, i.e. in the number of useful photons. It should thus be possible to satisfy the starting condition again with a relatively small increase in X/p, or an increase in potential of about 30 volts. An increase of about this value is in fact found: the accuracy of the measurements does not justify a more definite statement, but the (V_s, pd) curve for the long vessel does lie about 50 volts above that for the short vessel. Experiment and theory are thus consistent.

6.4. *The Slow Secondary Mechanism*

The essential property of the mechanism required is that it must be largely independent of time over a wide range. It must produce near the cathode wall at least one electron after a time interval ranging from 10^{-5} sec (corresponding to $\lambda=6000$ m) to 10^{-2} sec (corresponding to $f=50$ c/s) and perhaps longer.

When the current pulse ceases all the products of the avalanche except the electrons remain in the gas. Ions, moving with their mobility in the field due to that applied plus space and wall charges, are swept out in 10^{-4} sec or less, when $p=27$ mm. Resonance photons and metastable atoms both diffuse outwards and take about 10^{-3} or 10^{-2} sec respectively to travel across a 2 cm vessel, and rather less at lower pressures. If any of these agencies were responsible for the secondary electrons we should expect a change in the regularity or starting potential of the pulses at some frequency between 10 kc/s and 50 c/s. None was found, although it was sought carefully down to frequencies of 10 c/s.

There remain only the wall charges. It appears that the many slow electrons which are deposited on the 'anode' wall at the end of a pulse, and which cannot leak away on the insulating surface, are easily pulled off when the field reverses. This takes place in relatively low fields ($<1000\,\mathrm{v\,cm^{-1}}$) and is quite different from normal field emission.

Further experiments were done to check these conclusions.

§ 7. EXPERIMENTS ON THE EFFECT OF WALL CHARGES

The first set of experiments was designed to estimate how long after the end of a current pulse the slow secondary process remained effective. A d.c. field was applied across the electrodes in series with R and slowly increased until a current pulse occurred. The field was left on for a time T, and then reversed by a mechanical switch. By synchronizing the time base of the oscilloscope to give a single sweep when the field was reversed we could see how soon after the reversal another current pulse occurred. The time T was varied. A large resistance in series with the source prevented any possible surges but delayed the attainment of the full voltage by about 50 μsec.

The results obtained were: (i) The starting potential measured was the same as when alternating fields were applied. This confirms our conclusions about the starting conditions. (ii) The slow secondary process remains effective for a period of minutes. The time T was increased up to one minute, and a current pulse was always observed within 100 μsec of reversing the field. Fluctuations between 50 and 100 μsec were probably due to the simultaneous manual operation of two switches. Only when T was increased to about 3 minutes did current pulses occur at random times after reversing the field.

A slight variation of the procedure was then adopted. The d.c. field was applied, and switched off immediately after a current pulse occurred. The electrodes were then left with no applied field for the time T, after which a reverse field of the same size was applied. The results were exactly the same. The large values of T obtained show that it is impossible that diffusing particles are responsible for starting the next pulse, and are compatible with its being due to wall charges which are known to stay on glass for many minutes. The effect is also independent of the presence of the applied field during the time T, a further evidence in favour of the wall charge theory.

The second set of experiments was designed to remove the wall charges. A d.c. field was applied, and switched off after a current pulse occurred, exactly as above. A high-frequency field from a Tesla coil was then applied across the electrodes for a few seconds, forming an intense plasma in the middle of the vessel. After a subsequent time T, during which no external field existed, a reverse d.c. field was applied. The results were: (i) When T is less than about 10 seconds, a current pulse occurred immediately (i.e. within about 100 μsec) upon applying the reverse d.c. field. This may have been due to a few electrons from the intense plasma remaining in the gas. (ii) When T is more than about 10 seconds a current pulse was never observed to start as soon as a reverse field was applied, but at quite random times afterwards. Now if the starting of these current pulses were due to any ionization products in the gas then the high-frequency discharge, which increases the amount of ionization in the gas, should have made their starting more probable. However, the high-frequency discharge has the opposite effect on wall charges: when the applied field is turned off the plasma will disperse in such a way as to neutralize any remaining field, namely that due to the wall charges. These experiments are therefore reasonably conclusive that wall charges act as a source of secondary electrons to start current pulses.

ACKNOWLEDGMENTS

I wish to thank the Research Council of British Electricity Authority for a grant to carry out this research, Dr. J. S. Forrest, Director, B.E.A. Research Laboratories, for helpful discussion and the loan of apparatus and Mr. Oakeshott of B.E.A. Research Laboratories for technical assistance. I particularly wish to thank Dr. A. von Engel for constant advice and encouragement, and Professor Lord Cherwell for extending to me the facilities of the Clarendon Laboratory.

REFERENCES

BROSE, H. L., 1925, *Phil. Mag.*, **50**, 536.
EMELÉUS, K. G., LUNT, R. W., and MEEK, C. A., 1936, *Proc. Roy. Soc.* A, **156**, 394.
FRANCIS, G., and VON ENGEL, A., 1953, *Phil. Trans. Roy. Soc.* A, **909**, 246, 143.
GILL, E. W. B., and VON ENGEL, A., 1949, *Proc. Roy. Soc.* A, **197**, 107.
HALE, D. H., 1939, *Phys. Rev.*, **55**, 815.
HARRIES, W. L., and VON ENGEL, A., 1954, *Proc. Roy. Soc.* A, **222**, 490.
LLEWELLYN JONES, F., and DAVIES, D. E., 1951, *Proc. Phys. Soc.* B, **64**, 519.
PENNING, F. M., and MOUBIS, J., 1935, *Physica*, **2**, 55.
WAINFAN, N., WALKER, W. C., and WEISSLER, G. L., 1953, *J. Appl. Phys.*, **24**, 1318.
WHIDDINGTON, R., and JONES, H., 1928, *Phil. Mag.*, **6**, 889.

Photoconductivity in Calcium Tungstate

By J. R. COOK
Department of Physics Applied to Medicine, The Middlesex Hospital, London W.1

Communicated by J. E. Roberts ; MS. received 25th October 1954

Abstract. In view of earlier contradictory published results on the photoconductivity of crystalline calcium tungstate, this effect has been examined using artificial single crystals and an intense ultra-violet source. Photoconductivity has been demonstrated, with a space charge build up of long relaxation time. This latter is probably linked with the long duration phosphorescence previously reported in this material. The evidence produced indicates that calcium tungstate is an intrinsic semiconductor with an activation energy of 2·1 ev. The photoconductive properties are in approximate agreement with the assumption of an exponential trap distribution if a variation of capture cross section with temperature is postulated.

§ 1. INTRODUCTION

RANDALL AND WILKINS (1945) have measured the conductivity produced in a compressed layer of calcium tungstate powder by ultra-violet radiation and found a current of approximately 2×10^{-13} ampere. Their measurements were limited to a single determination of current. On the other hand, Klick and Schulman (1949) failed to find a photoconductive effect in a small artificially grown crystal of calcium tungstate illuminated by ultra-violet light from mercury and hydrogen discharge lamps. The maximum sensitivity of their electrometer system was 5×10^{-15} ampere.

In view of the difference between these experiments it was decided to recheck the effect with an artificial calcium tungstate crystal (Linde Air Co., America) and an intense light source.

§ 2. EXPERIMENTAL METHOD

Three cylindrical clear crystals have been used in this work, each of diameter 5 mm and length 3 mm. The crystal faces were ground so as to be perpendicular to the long axis, and were coated with a thick layer of colloidal graphite. Electrical contact on the two faces was effected by two brass contacts mounted on distrene insulators (figure 1). The thick carbon layer prevented possible photoelectric emission from the brass contacts, when the whole crystal was illuminated. The source of ultra-violet was a high pressure mercury lamp fitted with a Wood's filter and emitting its main line at 3650 Å.

A negative voltage was applied to one electrode, the other being connected to the current measuring device. This was either a Lindemann electrometer or a sensitive d.c. amplifier with a $10^9\ \Omega$ input resistance. In general the electrometer was used for experiments lasting several hours, because of its inherent stability. The minimum measurable current was approximately 10^{-13} ampere. The Lindemann electrometer was used in conjunction with a compensating circuit connected to the air condenser C so that the voltage on the collection electrode was kept constant. Identical results were obtained from the two methods.

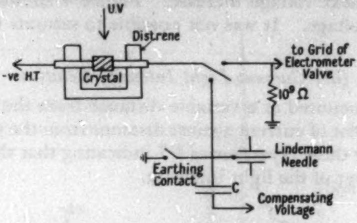


Figure 1. Experimental arrangement.

§ 3. EXPERIMENTAL RESULTS

In the case of three crystals we have obtained measurable photocurrents. The dark current at room temperature was always negligible compared with the photocurrent.

(i) *Space Charge Limited Photocurrents*

Figure 2 indicates the reduction in photocurrent as a function of time measured by the Lindemann electrometer over a period of three weeks under constant ultra-violet exposure. After three weeks the current has reached an approximately constant value. If the light source is cut off the crystal recovers over a period of several hours as indicated. Since the recovery is incomplete after 4000 minutes a permanent alteration of the crystal structure may have occurred. Figure 3 shows the variation in decay time of the initial region as a function of applied voltage. In order that current measurements should not be affected by space charge build-up it was essential to take photocurrent readings within one minute of exposing the crystal to the source and then to cover the source.

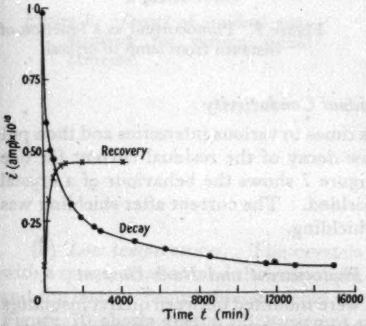


Figure 2. Space charge decay under constant light intensity and recovery measured by intermittent illumination.

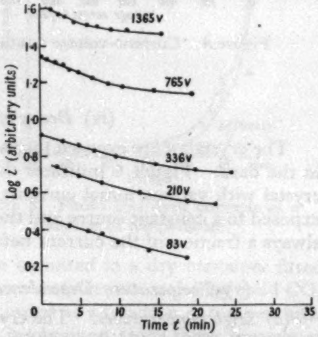


Figure 3. Decay of photocurrent under constant illumination.

(ii) *Current–Voltage Relation*

The crystals were exposed to the source and the current measured and then shielded before the next voltage increase. Figure 4 shows that the current is proportional to the voltage. It was not possible to saturate the current with the voltages available.

(iii) *Current Light Intensity Relation*

The source was mounted at a variable distance from the crystal and figure 5 shows a logarithmic plot of current against distance from the source to the crystal. The average slope for these crystals was 1·7 indicating that the current is proportional to the 0·85 power of the light intensity.

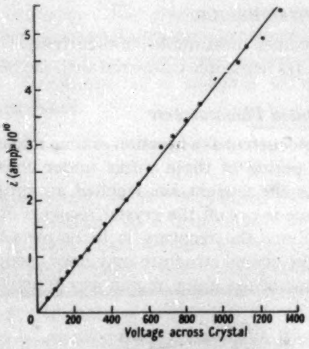


Figure 4. Current–voltage relation.

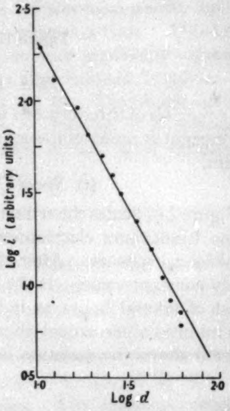


Figure 5. Photocurrent as a function of distance from lamp to crystal.

(iv) *Decay of Residual Conductivity*

The crystals were exposed for various times to various intensities and then put in the dark. Figure 6 indicates the slow decay of the residual current for one crystal with various initial currents. Figure 7 shows the behaviour of a crystal exposed to a constant source and then shielded. The current after shielding was always a fraction of the current before shielding.

(v) *Temperature Dependence of Photocurrent and Dark Current*

(*a*) *High temperatures*. The crystals were mounted between quartz insulators in an oven provided with a window so the photocurrent could be measured as a function of temperature. As the temperature was raised above 80°c in all cases the dark current rapidly increased (figure 8). If the crystal was illuminated above 80°c the current increased from the photocurrent process. Before the onset of dark current the photocurrent was itself found to increase with temperature. When the dark current became measurable (around 80°c) the photocurrent became a constant perturbation on the dark current. Measurements of dark current were extended to 170°c with the d.c. amplifier. Figure 9 shows that a plot of log dark current as a function of the reciprocal temperature (°K) is accurately linear. Similar curves to those shown in figures 8 and 9 were obtained for the three crystals.

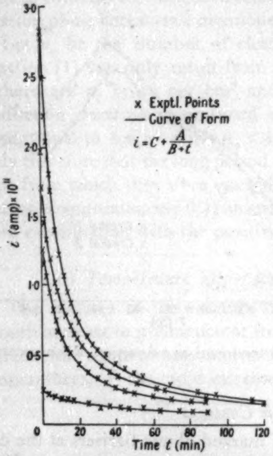


Figure 6. Decay of residual conductivity for various initial currents.

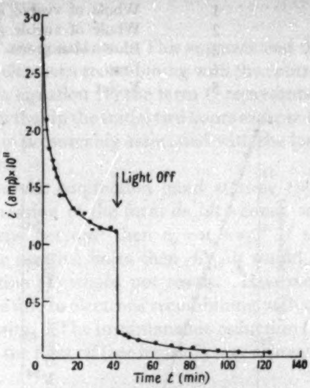


Figure 7. Space charge effect and recombination current.

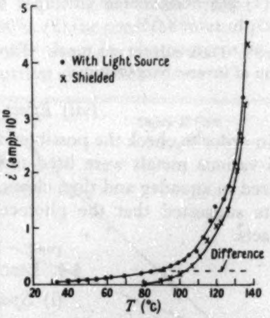


Figure 8. Photo and dark currents as a function of temperature.

(*b*) *Low temperatures*. The crystals were mounted in a dry container fitted with a quartz window and thermocouple. The container was cooled by solid CO_2 and the instantaneous photocurrents were recorded as a function of temperature. Figure 10 shows that a photocurrent peak is obtained from these crystals at -15°c.

(vi) *Photoconductivity from Visible Light*

A small current was seen to occur in the crystals when exposed to daylight. The crystal was therefore exposed to the intense illumination of a white light and a series of Wratten filters was interposed between it and the lamp. An increase of photocurrent occurred from the red to the violet end of the spectrum as follows.

The currents in this table are corrected for equal intensities of transmitted light as determined with a photocell of known response.

Filter	Filter transmission	Relative photocurrent
1	Whole of visible red	0
2	Whole of visible green	8
3	Blue + blue green	170
4	Blue	220
5	Violet	710

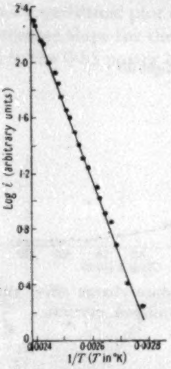


Figure 9. Dark current as a function of inverse temperature.

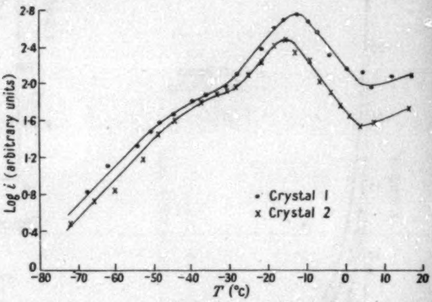


Figure 10. Photocurrent as a function of temperature.

(vii) *Effect of Electrode Contacts*

In order to check the possible existence of marked ohmic barriers at the contacts various metals were used to touch the crystal, whose surfaces were both covered in aquadag and then cleaned. Results identical to all the above experiments suggested that the photocurrent was not appreciably affected by the contacts.

§ 4. DISCUSSION OF RESULTS

(i) *Space Charge Effects*

The curves obtained, of which figure 2 is an example, were very similar to those of von Hippel *et al.* (1953) in alkali halides crystals containing F centres. In particular the approximately exponential decrease of current with time in the initial part of the curves at low voltage (figure 3) and the tendency to a constant final value of current agree with their results qualitatively. It was not considered worth while to attempt a quantitative fit to their theory but two points may be noted. The basic assumption that the electrons leave a positive space charge behind must apply in the present case, and the steady final current may be associated with field emission. The long period relaxation time may be associated with trapping effects. It is then unnecessary to postulate a very low electron drift mobility. In this connection the experiments of Gillette (1950) on a long duration phosphorescence in such a crystal are of interest in that they indicate the presence of traps which give the electron a very small escape probability.

(ii) *Decay of Residual Conductivity*

The decay curves shown in figure 6 fit formulae of the type

$$i = C + \frac{A}{B+t} \qquad \ldots\ldots(1)$$

where i is the current at time t, and A, B, C are constants. This suggests that the crystal is returning to its initial state by the electrons recombining with the centres from which they were initially ejected. In equation (1) the term C represents a current whose rate of decay is much less than that in the initial two hours and can be taken as constant for the curve fitting. It is presumably associated with the long duration phosphorescence mentioned above.

Let n_c be the number of electrons in the conduction band at time $t=0$. Equation (1) can only result from an expression of the form $dn_c/dt = \text{const. } n_c^2$. If there are n_p holes per cm³ and n_t traps per cm³ then $n_p = n_c + n_t$. If the conduction electrons recombined with the positive holes then dN_c/dt would be proportional to $n_c n_p$. Since $n_c < n_t$ equation (1) would not result. It is more likely therefore that the long period decay is due to electrons recombining with the traps from which they were ejected thermally. The instantaneous reduction (by a factor of approximately 0·1) on switching the light off (see figure 7) may represent rapid combination with the positive holes.

(iii) *Temperature Dependence of Photocurrent and Dark Current*

The features of temperature dependence requiring explanation are (1) the smooth increase in photocurrent from 0°c to 80°c, (2) the constant value of photocurrent for temperatures above 100°c, (3) the drop in photocurrent at low temperatures, (4) the rapid increase in dark current above 80°c.

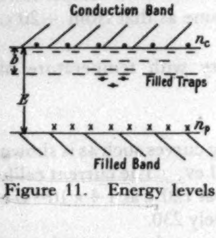


Figure 11. Energy levels.

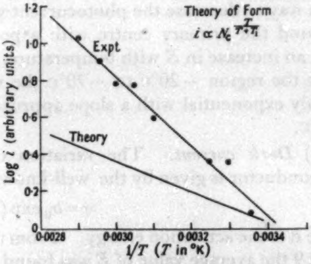


Figure 12. Log photocurrent as a function of reciprocal temperature for one crystal. $T_1 = 1660$°K.

(*a*) *Photocurrent*. The fact that the current varies as the 0·85 power of the light intensity suggests that the type of model proposed by Rose (1951) may be applied. For a crystal having an exponential distribution of traps below the conduction band (see figure 11), Rose deduced the equation

$$n_c = \left(\frac{f}{V S A T N_c^{-T/T_1}} \right)^{T_1/(T+T_1)}$$

where f is the number of optical excitation sec^{-1} cm^{-3}, V the thermal velocity of electrons (cm sec^{-1}), S the capture cross section (cm^2) of the primary centre for free carrier, N_c the number of energy levels in the bottom slice of the conduction band a few times $\boldsymbol{k}T$ wide, T the absolute crystal temperature, T_1 the temperature at which the traps were 'frozen in' and A a constant. The trap distribution is represented by

$$N_b\, db = A \exp(-b/T_1)\, db$$

where b is the depth of the energy level below the conduction band.

Since in the present experiments n_c is proportional to $f^{0\cdot85}$, T_1 must be approximately 1660°K. Since the melting point of a calcium tungstate crystal is 1500°C it is reasonable to suppose that the traps are 'frozen-in' during solidification with the given energy distribution.

If the product VS remains constant the variation of n_c with T is given by $n_c = \text{const.}\, N_c^{T/(T+T_1)}$ since $1/T$ does not vary appreciably. With $N_c \simeq 10^{19}$ cm^{-3}, the variation of current with T can be found. Figure 12 shows the comparison between this theory and experiment for one crystal in the range 20°C to 80°C. The theory gives an approximately exponential increase of n_c with T but the rate of increase differs from the experiment, being smaller by a factor of three.

At temperatures above 80°C the dark current increases rapidly. The number of conduction electrons will then approach the number of trapped electrons and ultimately trapping effects will be negligible. This means that when the crystal is illuminated an additional current should occur which will remain constant as the temperature is increased. This appears to be the case in figure 8.

In the above treatment the product VS was taken as constant. If however the capture cross section varies with temperature the discrepancy between theory and experiment may partly be explained. Such a temperature dependence is suggested by the peak obtained in figure 10 at -15°C. Rose (1951) has suggested that a way to decrease the photocurrent with increase in temperature would be to surround the primary centre with a potential barrier. This would effectively cause an increase in S with temperature.

In the region -20°C to -70°C the variation of current with temperature is roughly exponential with a slope approximately the same as that from $+20$°C to $+70$°C.

(*b*) *Dark current.* The variation of conductivity with temperature of a semiconductor is given by the well-known equation

$$\sigma = \sigma_0 \exp(-E/2\boldsymbol{k}T)$$

where E is the activation energy. From the slope of log curves such as is shown in figure 9 the average value of E was found to be $2\cdot1 \pm 0\cdot1$ ev. The current calibration of the d.c. amplifier gave the resistivity of a crystal at 120°C as $1\cdot4 \times 10^{11}\ \Omega$ cm. Using the above relation σ_0 was found to be approximately 230.

Moss (1952) has shown that the value of σ_0 differentiates between an impurity semiconductor and one having intrinsic conductivity. A value of σ_0 of the order of 230 and the large value of E suggests that at least the dark conductivity is intrinsic. In addition, the large value of E suggests the absence of impurity, which in general does not give rise to deep trap levels.

Since a value of 2·1 ev corresponds to a wavelength of 5850 Å it is to be expected that the photocurrent should increase from the red end of the visible spectrum as shown in the table. Moss (1952) has suggested that the variation in sensitivity of a photoconductor with wavelength is associated with an energy distribution of levels from which the photoelectrons originate. This fact could explain the variation of sensitivity with wavelength seen in the table.

§ 5. CONCLUSION

The present experiments show that calcium tungstate is an intrinsic semiconductor with an activation energy of 2·1 ev. The photoconductive properties suggest that an exponential trap distribution may be applicable if a variation ot capture cross section is allowed. The decay of photocurrent after the source is cut off is accurately binomial. In view of the long relaxation time for space charge build up it is not surprising that long duration phosphorescence can be observed.

ACKNOWLEDGMENTS

The author would like to thank Professor J. E. Roberts for helpful criticism and Mr. K. A. Mahmoud for assistance during the initial stage of this work.

REFERENCES

GILLETTE, R. H., 1950, *Rev. Sci. Instrum.*, **21,** 294.
VON HIPPEL, A., GROSS, E. P., JELATIS, J. G., and GELLER, M., 1953, *Phys. Rev.*, **91,** 568.
KLICK, C. C., and SCHOLMAN, J. H., 1949, *Phys. Rev.*, **75,** 1606.
MOSS, T. S., 1952, *Photoconductivity in the Elements* (London: Butterworths Scientific Publications).
RANDALL, J. T., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc.* A, **184,** 347.
ROSE, A., 1951, *R.C.A. Rev.*, **12,** 362.

Two New Simplified Systems for the Correction of Spherical Aberration in Electron Lenses*

By G. D. ARCHARD
Associated Electrical Industries Limited, Aldermaston Court, Aldermaston, Berkshire

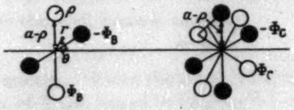
MS. received 23rd July 1954 and in final form 8th November 1954

Abstract. Two new simplified electrostatic systems for the correction of third-order spherical aberration in electron lenses are described and related to earlier systems of Seeliger and Burfoot. Numerical examples are given in which the voltages and positioning tolerances of the various electrodes are calculated. The effects of scaling and change of electrode shape are discussed.

---

§ 1. INTRODUCTION

ELECTRODE systems for the correction of spherical aberration in electron lenses have been proposed by Seeliger (1949, 1951), following mathematical analysis of Scherzer (1947), and by Burfoot (1952). Both systems are somewhat complicated. Seeliger's contains a large number of electrodes (two cylindrical einzel lenses, three groups of eight electrodes, and an auxiliary round einzel lens in addition to the lens under correction). Burfoot's system contains essentially only four electrodes, but these would require very complicated shapes and be limited by stringen tolerances. A modification of Seeliger's system has been suggested (Archard 1954, to be referred to as I) in which cylindrical einzel lenses would be replaced by 'Φ_2 units' (figure 1(*a*)). The virtue of this replacement lies in the fact that 'Φ_2 units' require much lower voltages (factor of 100) in order to produce the same cylindrical lens action. The modified system would nevertheless require the presence of two auxiliary round einzel lenses.



Φ_2 Unit (*a*) Φ_4 Unit (*b*)

Figure 1. Electrode groups of four- and eightfold symmetry.

In the following, two alternative simplifications are envisaged. In both cases, auxiliary round and cylindrical lenses are completely eliminated, and the number of electrodes is reduced by superimposing the 'Φ_2 units' on the 'Φ_4 units' (groups of eight electrodes, figure 1(*b*)) of Seeliger's system.

---

* One of these was briefly described in the course of a paper read at the International Conference on Electron Microscopy, London, July 1954.

§ 2. Basic Requirements of a Correction System

Scherzer's theoretical treatment (1947) shows that, in order to achieve spherical correction, a cylindrical lens action is needed, i.e. rays leaving the axial point of the object at unit inclination and travelling in the xz and yz planes (these rays will be called x_α, y_α respectively) must be differently deflected; potential systems with fourfold symmetry (Φ_4) must then be established at points, or strictly in limited regions, where x_α and y_α are at different distances from the axis. Finally, in order to prevent distortion, x_α and y_α must be brought together again by further cylindrical lens action. These conditions are summarized in Scherzer's equations (1.4) and (4.5), the former being the ray equations for x_α and y_α, which must emerge from the system superimposed, and the latter giving the condition for the correction of spherical aberration, from which it may be shown that no correction can be achieved unless Φ_4 potential forms exist where x_α^2 and y_α^2 are unequal.

Correction is achieved in Seeliger's system (figure 5(*a*)) by separately focusing x_α and y_α (in the two-dimensional diagram, the rays so marked must be thought of as travelling in the xz, yz planes respectively), and at each line focus establishing a 'Φ_4 unit'. This achieves the correction of spherical aberration in two planes at right angles but, unless a certain impracticable relation (which will be mentioned later) exists between the Φ_2 and Φ_4 voltages, leaves some residual spherical aberration in the two planes at 45° to the former. Insertion of a third Φ_4 unit in a region where x_α and y_α are (schematically) superimposed completes the correction for all planes. Distortion does not occur because, by virtue of the symmetry of the system, x_α and y_α emerge at the same distance from the axis and at the same inclination thereto.

In Burfoot's system (figure 5(*c*)), separate line focusing is dispensed with. His first electrode acts as the combination of a diverging round lens and a Φ_2 unit, achieving the initial separation of x_α and y_α. His second electrode acts as the combination of a converging round lens and two Φ_4 units, and his third electrode has properties similar to the first. At a subsequent point, x_α and y_α again attain equal distances from the axis, and are then differentially deflected by the fourth electrode (a Φ_2 unit) so as to assume the same inclination to the axis, thus eliminating distortion.

§ 3. Description of New Simplified System, Type A

Consider figure 2, which relates to the correction of an electron microscope objective lens. The rays marked x_α, y_α leave the object point O at inclination unity, pass through the objective lens at height f (focal length), and emerge almost parallel to the axis. More precisely, their emergent inclination is the reciprocal of the uncorrected magnification. A short distance from the objective, the rays enter a Φ_2 unit; one acquires a smaller slope, the other a greater. At the point where y_α crosses the axis, a Φ_4 unit (figure 1(*b*)) is placed. This does not affect the course of x_α, y_α as it does not enter into the equations which guide them (Scherzer 1947, eqn (1.4)). At the point where x_α crosses the axis, two further units are placed, a Φ_2 and a Φ_4. In practice, potentials pertaining to the two kinds would be superimposed on the same electrodes. The Φ_2 component here is stronger than the previous Φ_2 unit, as it has to turn y_α back again towards the axis; x_α, because it passes through the centre of the unit, is undiverted. At the point where x_α and y_α again meet, a third Φ_2 unit reunites them.

This system resembles Burfoot's rather than Seeliger's as it employs essentially two Φ_4 and three Φ_2 potential forms. Burfoot demonstrated mathematically the necessity for a third Φ_2 unit in order to avoid first-order distortion. This is another example of the well-known principle (e.g. Scherzer 1947) that three, but not two, cylindrical lenses may be made to simulate the properties of a round lens without distortion. (In Seeliger's system (figure 5(*a*)), the action of a third Φ_2 unit is effectively present, concealed in the central round lens, for the x_α and y_α rays enter this on opposite sides of the axis, and hence are deviated in opposing senses.)

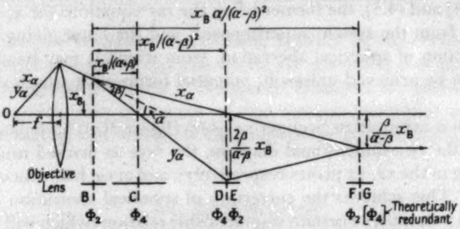


Figure 2. Simplified spherical correction, system A.

§ 4. VOLTAGES OF SYSTEM A

Let the Φ_2 and Φ_4 units (figure 1) comprise groups of spheres of radius ρ with centres distant *a* from the axis, alternate spheres bearing potentials $\pm\Phi_B$, $\pm\Phi_C\ldots$ Let Φ_A be the accelerating potential and (x_B, y_B), $(x_C, y_C)\ldots$ indicate the values of (x_α, y_α) at the centres of units B, C…(figure 2). Then Scherzer's (1947) equations (4.5) may be modified as shown in I, equations (1), (2), (3), (15), (16). The result is:

$$C_s+6\cdot45(\rho^2/a^5)(x_B{}^4\Phi_B{}^2/\Phi_A{}^2+x_F{}^4\Phi_F{}^2/\Phi_A{}^2)$$
$$-8(\rho/a^4)(x_C{}^4\Phi_C/\Phi_A+x_G{}^4\Phi_G/\Phi_A)=0 \qquad \ldots\ldots(1)$$

$$C_s+6\cdot45(\rho^2/a^5)(y_B{}^4\Phi_B{}^2/\Phi_A{}^2+y_E{}^4\Phi_E{}^2/\Phi_A{}^2+y_F{}^4\Phi_F{}^2/\Phi_A{}^2)$$
$$-8(\rho/a^4)(y_D{}^4\Phi_D/\Phi_A+y_G{}^4\Phi_G/\Phi_A)=0 \qquad \ldots\ldots(2)$$

$$C_s-3\cdot86(\rho^2/a^5)(x_B{}^2y_B{}^2\Phi_B{}^2/\Phi_A{}^2+x_F{}^2y_F{}^2\Phi_F{}^2/\Phi_A{}^2)$$
$$+24(\rho/a^4)x_G{}^2y_G{}^2\Phi_G/\Phi_A=0. \qquad \ldots\ldots(3)$$

C_s represents the spherical aberration constant of the lens under correction, and is essentially positive—a typical value would be 0·5 cm. (C_s in the case of (I) included contributions from two other round lenses contained in the system (figure 5(*b*)).) It will be seen from (3) that the element G (figure 2) is, in principle, unnecessary, for (3) could be satisfied by choosing suitable values of Φ_B, Φ_F, and then (1), (2) will be satisfied by duly adjusting Φ_C, Φ_D. But, as units B and F also enter into first order calculations and have to be adjusted so as to re-synthesize rays x_α and y_α, it is convenient to have an extra variable available. Its omission will be found to lead to inconveniently high voltages on B and F.

A typical case will now be considered numerically. Figure 2 shows that, for small angles α, β, and for rays finally leaving the system parallel to the axis, the total length of the correction system is $x_B\alpha/(\alpha-\beta)^2$. x_B may be expected to be of the order *f*, typically 0·5 cm.

Consider a ray having $\alpha = 0{\cdot}2$, $\beta = 0{\cdot}1$. Then the length of the correction system will be 10 cm. It is shown elsewhere (Archard 1955, to be referred to as II) that the change of slope produced by a Φ_2 unit (say B) is $4(\rho/a)(x_B/a)\Phi_B/\Phi_A$.

Suppose $\rho = 0{\cdot}05$ cm, $a = 0{\cdot}25$ cm. To produce a deflection $\beta = 0{\cdot}1$, Φ_B/Φ_A must be 0·0625, i.e. for an accelerating potential 50 kv, Φ_B must be 3 kv.

Likewise, for E to produce a deflection 2α, Φ_E must be 6 kv, and, for F to produce a deflection $\alpha - \beta$, Φ_F must be 3 kv. The 6 kv would be rather inconvenient, as the gaps between adjacent spheres (at $\pm\Phi_B \ldots$) would be only 0·25 cm. The voltages can, however, be reduced (as detailed in II) by substituting groups of laminae for groups of spheres. For the same value of a, laminae with opposite edges separated by $2a$ and extending a distance l along the axis show a voltage reduction factor of $l/2\pi\rho$. Thus, for $l = 1$ cm, the above voltages would be reduced by a factor of three, leaving $\Phi_B = \Phi_F = 1000$ v, $\Phi_E = 2000$ v, and the interelectrode gaps would be increased to nearly 0·35 cm. If the value of a were now halved, the voltages would fall by a further factor of four while the interelectrode distances would fall by a factor of two. On the other hand, as will appear later, tolerances would worsen by a factor of sixteen, so that it is in the best interest to keep the voltages as high as the insulation will stand.

If now these voltages are substituted into (1), (2) and (3) (adjusted to allow for the change in electrode shape), it transpires that terms containing Φ_B, Φ_F are very small, and that containing Φ_E fairly small, compared with the typical C_s value of 0·5 cm. It would in fact require an extremely strong Φ_2 system to satisfy (3) without the presence of the Φ_4 unit G, and this must therefore be regarded as a practical necessity, though algebraically redundant.

Under these circumstances, equations (1), (2) and (3) may be reduced to

$$C_s - 8(\rho/a^4)x_C{}^4\Phi_C/\Phi_A - 8(\rho/a^4)x_G{}^4\Phi_G/\Phi_A = 0 \qquad \ldots\ldots(4)$$

$$C_s - 8(\rho/a^4)y_D{}^4\Phi_D/\Phi_A - 8(\rho/a^4)y_G{}^4\Phi_G/\Phi_A = 0 \qquad \ldots\ldots(5)$$

$$C_s + 24(\rho/a^4)x_G{}^2y_G{}^2\Phi_G/\Phi_A = 0. \qquad \ldots\ldots(6)$$

Now $x_G = y_G = \{\beta/(\alpha-\beta)\}x_B \to x_B$ in the present case; hence from (6),

$$\Phi_G/\Phi_A = -0{\cdot}026,$$

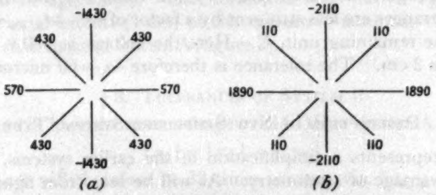


Figure 3. Superimposition of Φ_2 and Φ_4 voltages.

i.e. $\Phi_G = 1300$ v. If, however, a group of eight laminae, extending 1 cm along the axis, is substituted for the eight spheres of unit G, it will be possible to superimpose G and F as indicated in figure 3(*a*). It is shown in II that this will result in a lowering of the required potential by a factor of $l/2\pi\rho$ (in complete analogy with the Φ_2 case), so that Φ_G will be reduced three times, and will become $\Phi_G = 430$ v. Substitution of (6) in (4) and (5) now gives

$$(4/3)C_s = 8(\rho/a^4)x_C{}^4\Phi_C/\Phi_A = 8(\rho/a^4)y_D{}^4\Phi_D/\Phi_A. \qquad \ldots\ldots(7)$$

Consider unit D: $y_D = \{2\beta/(\alpha-\beta)\}x_B \rightarrow 2x_B$. Thus $\Phi_D/\Phi_A = 0{\cdot}0065$ giving (for spheres) $\Phi_D = 325$ v.

Again it will be convenient to replace spheres by laminae so as to facilitate the superposition of D and E. Consequently $\Phi_D = 110$ v.

Consider unit C: $x_C = \{2\beta/(\alpha+\beta)\}x_B \rightarrow 2x_B/3$. Thus $\Phi_C/\Phi_A = 0{\cdot}53$. This represents ± 27 kv across a gap of 1 mm, which is quite impracticable. But the use of eight laminae, this time extending 2 cm along the axis and having opposite edges 0·3 cm apart, would reduce this voltage by a factor of 46 so that there would exist a potential difference of 1160 v (i.e. ± 580) across a gap which would still be 1 mm, and this would not be impracticable. In this case, therefore, $\Phi_C = 580$ v. The horizontal lines above B, C ... in figure 2 represent the axial extents of the respective laminae, from which it is seen that C could not be made to extend more than 2 cm parallel to the axis.

§ 5. TOLERANCES OF SYSTEM A

It is shown in I that slight radial or rotational displacement of one of the electrodes of a four or eight sphere unit introduces a 'Φ_1' action which produces second and fourth order aberrations. The second order can be removed if arrangements are made to impart slight antisymmetrical changes to the potentials of opposite spheres of the Φ_4 units. The fourth order aberrations do not exceed $C_s\theta^3/n$ (i.e. $1/n$ of the radius of the disc of spherical aberration, θ being the semi-aperture angle of the beam, typically about 0·005 radians), if

$$\delta a < (C_s/\theta n)(\Phi_A/\Phi_B)(a^6/20\rho r^5) \qquad \ldots\ldots(8)$$

where $r = x_\alpha$ or y_α at the unit in question. The corresponding calculations for laminar electrodes are intractable, but it seems likely that an idea of their tolerances will be gained by substituting $l = 2\pi\rho$ in (8).

Consider electrode group DE, whose highest voltage is 2110 v, $r = 2f$, $a = 0{\cdot}25$, $l = 1$ cm. In order to secure a five times improvement in the resolution as limited by spherical aberration and diffraction, n must be 5^4. Then $\delta a_{DE} = 2{\cdot}8$ microns.

Electrode groups B and FG have $r = f$ and voltage of the order of 1000 v; hence, their tolerances are less stringent by a factor of 64: $\delta a_{B,\,FG} \sim 180$ microns.

Consider the remaining unit, C. Here the voltage is 580 v, r is $2f/3$, a is 0·15 cm and l is 2 cm. The tolerance is therefore $\delta a_C = 60$ microns.

§ 6. DESCRIPTION OF NEW SIMPLIFIED SYSTEM, TYPE B

System A represents a simplification of the earlier systems, but it suffers from the disadvantage of asymmetry. As will be seen from figure 2, electrode groups B and C come relatively close together, and the voltages required for C are disproportionately large. At the expense of the addition of one extra Φ_2 unit, however, the asymmetry can be removed and voltages distributed more evenly, while the actual number of electrodes remains the same.

Consider figure 4. Rays x_α and y_α leave the objective lens parallel to the axis and at a distance $x_L = f$ from it. They enter a Φ_2 unit and become separated. At the point where y_α crosses the axis, another Φ_2 is placed; this turns x_α back so as to run parallel to y_α. At the point where x_α crosses the axis, a similar Φ_2 unit is placed, and the rays are re-synthesized symmetrically. Φ_4 units are conveniently superimposed on the second and third Φ_2 units. The third Φ_4 unit may be placed as shown in the centre of symmetry, or superimposed on the first or fourth Φ_2 unit.

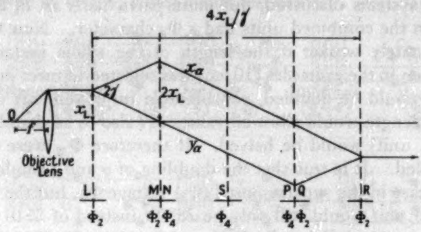


Figure 4. Simplified spherical correction, system B.

§ 7. VOLTAGES OF SYSTEM B

If unit L deviates the rays through an angle $\pm\gamma$, the total length of the system will be $4x_L/\gamma$. Hence, in order to keep the same total length as system A (10 cm),

$$4/\gamma = \alpha/(\alpha-\beta)^2 \qquad \ldots\ldots(9)$$

or in the particular case considered $\gamma = \alpha = 2\beta = 0{\cdot}2$ radian.

Thus, L and R will have the same voltage as E (which caused a deviation of 2α but had $y_E = 2x_L$). For laminar electrodes having $a = 0{\cdot}25$ cm and $l = 1$ cm, $\Phi_L = \Phi_R = 2000$ v.

Units M and P, like E, cause a deviation 2γ, but have $|x_M| = |y_P| = 2x_L$. Consequently, they bear the same voltage as L and R: $\Phi_M = \Phi_P = 2000$ v.

Whether Φ_S is placed as shown or superimposed on L or R, $x_S{}^2 y_S{}^2$ will be exactly the same as $x_G{}^2 y_G{}^2$ in system A. Hence, for laminae having $a = 0{\cdot}25$ cm, $l = 1$ cm, $\Phi_S = 430$ v.

The remaining Φ_4 units N and Q have the same values of $x_\alpha{}^2$, $y_\alpha{}^2$ as D in system A. For similar laminae, therefore, $\Phi_N = \Phi_Q = 110$ v. The combination MN is represented in figure 3(*b*). The horizontal lines above Φ_2, Φ_4 in figure 4 represent the axial extents (1 cm) of the respective laminae. The laminae are better spaced than those of system A (figure 2).

§ 8. TOLERANCES OF SYSTEM B

Units L and R have the same voltage as did DE in system A, but

$$|x_L| = |x_R| = \tfrac{1}{2}|y_E|.$$

Their tolerances will therefore be 32 times greater, viz. $\delta a_{L,\,R} = 90$ microns for a five times improvement in resolution.

Units MN and PQ, however, have the same voltage as L and R but $|x_M| = |y_P| = 2|x_L| = |y_E|$. Hence $\delta a_{MN,\,PQ} = 2{\cdot}8$ microns. In its separated state, unit S will have $\delta a_S = 450$ microns.

§ 9. SCALING

Consider a laminar Φ_4 unit of constant strength (e.g. $\Phi_S l/a^4 =$ constant). For the moment, let l remain constant. Then $\Phi_S \propto a^4$, but electrode spacing $\propto a$; hence $\Phi_S/\text{spacing} \propto a^3$; that is, insulation is improved by reducing a. On the other hand, tolerances $\propto a^6$/voltage, i.e. $\propto a^2$. Hence, it is in the best interests to keep the voltages as high as the insulation will allow. This is still more important for a Φ_2 unit, where $\Phi \propto a^2$, Φ_s/spacing $\propto a$, but tolerances $\propto a^4$.

In both the systems discussed, but more particularly in B, the greater part of the voltage on the combined units had a Φ_2 character. Now Φ_2 units can be made proportionately weaker if the length of the whole system is increased. The length chosen in the examples (10 cm) was selected to meet practical requirements but, if it could be doubled, considerable improvements could be made. Doubling of the length would allow all values of l also to be doubled, at the same time $\Phi_M l/a^2$ (Φ_2 unit) would be halved. If therefore Φ_M were kept the same, a could be doubled. It is true that the doubling of a and l would necessitate an eight times increase in the superimposed (Φ_4) voltage Φ_N, but the highest voltage on the combined unit would still only be 2880 v instead of 2110 v. Hence, the tolerance of the lamina bearing that voltage would be $2^5 . 2110/2880 = 23$ times its previous value, i.e. $\delta a_{MN,\,PQ} = 66$ microns. It is plain that such an increase in tolerance would easily accommodate normal workshop practice.

§ 10. COMPARISON OF THE VARIOUS SYSTEMS

A schematic comparison of the various systems is given in figure 5. In order to pass from Seeliger's system (figure 5(*a*)) to the modification proposed in I (figure 5(*b*)), the cylindrical einzel lenses are first replaced by round lenses plus Φ_2 units, then one of the round lenses is absorbed into the objective, and another forms an intermediate stage or helps with the projection. In passing from this to Burfoot's system (figure 5(*c*)), the three Φ_4 units move inwards, degenerate into two Φ_4-type electrode-shaping parameters, link with a round lens component, and form the complicated shape of Burfoot's second electrode. The Φ_2 action latent in Seeliger's central lens transfers to Burfoot's fourth electrode, which reassembles the x_α and y_α.

System A (figure 5(*d*)) comprises, in effect, Burfoot's system without its round lens components, all the ' round ' potential components having withdrawn into the objective. Burfoot's first electrode becomes merely a Φ_2 unit, the two Φ_4 shaping parameters of his second electrode re-separate into Φ_4 units, but one of them immediately links with the residue of his third electrode, viz. a Φ_2 unit. The fourth electrode is unchanged, but for reasons which have been indicated an extra Φ_4 unit may be superimposed upon it.

The combined units are represented pictorially as four large circles (representing the Φ_2) concentric with four of a group of eight smaller circles (representing the Φ_4). The unshaded parts then signify positive voltages and the shaded parts negative (cf. figure 3).

In system B (figure 5(*e*)), the asymmetry of system A has been eliminated by a redistribution of the Φ_2 action. Schematically speaking, the Φ_2 unit E splits up into two Φ_2 units, one of which links with the Φ_4 unit C and forms MN in the new system, and the other of which links with G and reappears as PQ. This leaves D (a Φ_4 unit) without any attached Φ_4, and it reappears in the centre of symmetry as S. The Φ_2 unit F (which has lost its attached Φ_4) now appears as R and has the function of reassembling the x_α and y_α rays in system B.

The change in location of the various potential components on moving from system to system is shown in the figure by the arrowed encircled symbols Φ, Φ_2 and Φ_4, of which the first represents the round lens components and the others represent the Φ_2 and Φ_4 components respectively.

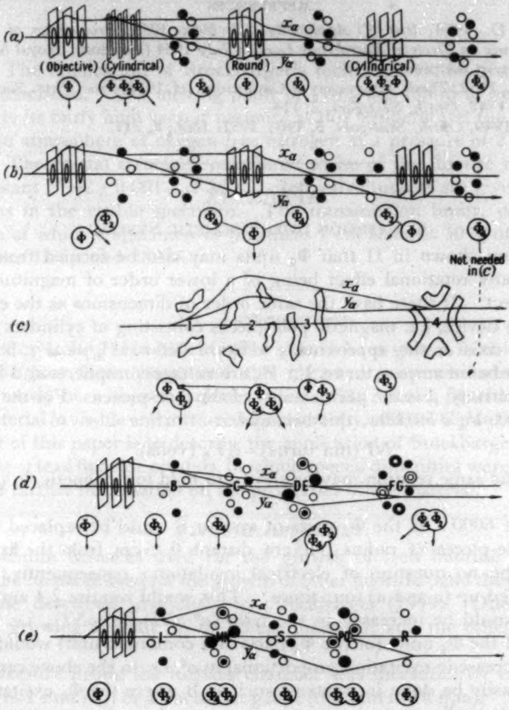


Figure 5. Relation between various correction systems. (*a*) Seeliger, (*b*) modification proposed in I, (*c*) Burfoot, (*d*) new system A, (*e*) new system B.

§ 11. CONCLUSIONS

Two systems for the correction of spherical aberration have been discussed; these eliminate the round and cylindrical einzel lenses and complicated electrode shapes involved in earlier suggestions. Examples have been given for systems 10 cm long, applied to the correction of lenses having $f=C_s=0{\cdot}5$ cm. By the use of laminar electrodes, voltages are kept below ± 2500 v, but positioning tolerances on some of the electrodes are rather stringent, e.g. for a five times improvement in resolution, 2·8 microns. Doubling the length of the systems would increase the lowest tolerance to 60 microns. These tolerances are subject to the provision of antisymmetrical fine-adjustment voltages to be applied to opposite electrodes of the 'Φ_4' units.

ACKNOWLEDGMENTS

The author wishes to thank Mr. M. E. Haine and Dr. G. Liebmann for helpful comment, and Dr. T. E. Allibone for permission to publish this paper.

REFERENCES

ARCHARD, G. D., 1954, *Brit. J. Appl. Phys.*, **5**, 294; 1955, *Proceedings of International Conference on Electron Microscopy London July* 1954 (London: Royal Microscopical Society), in the press.

BURFOOT, J. C., 1952, *Thesis*, University of Cambridge (cf. 1953, *Proc. Phys. Soc.* B, **66**, 775).

SCHERZER, O., 1947, *Optik, Stuttgart*, **2**, 114.

SEELIGER, R., 1949, *Optik, Stuttgart*, **5**, 490; 1951, *Ibid.*, **8**, 311.

APPENDIX

CONVERSION INTO MAGNETIC SYSTEM

It has been shown in II that Φ_2 units may also be formed from magnetic pole-pieces, any rotational effect being of a lower order of magnitude than the deflective effect. If these have the same order of dimensions as the electrostatic (four-sphere) device, i.e. magnetic pole-pieces consisting of cylinders of radius ρ distant from axis a, the approximate relation $NI = 11V_S/\mu\sqrt{V_A}$ holds, where NI is the number of ampere turns, V_S, V_A are voltages on spheres and accelerating voltage respectively, μ is the permeability of the pole-pieces. For the reasonable values $\mu = 500$, $V_A = 50\,000$ v, this becomes

$$NI\ (\text{mA turns}) = \tfrac{1}{10}V_S\ (\text{volts}).$$

Exactly the same relation may be shown to hold for magnetic Φ_4 (eight-pole) units.

Thus, the 6000 v on the Φ_2 units of system B could be replaced by 600 mA turns on pole-pieces of radius 0·05 cm distant 0·25 cm from the axis. There would now be no question of electrical insulation; consequently, the whole could be scaled up (ρ and a) four times. This would require 2·4 ampere turns. Tolerances would be increased in the ratio $(a^6/\rho NI)_1/(a^6/\rho NI)_2$, i.e. 250 times. Scaling up of the Φ_4 units (or the Φ_4 parts of a combined unit) would, however, require an increase in excitation proportional to a^3, i.e. in the above case 64 times. This could easily be done for systems such as B where the Φ_4 excitation before such scaling is small.

Growth of Lead Fluoride Crystals from the Melt

By D. A. JONES

Department of Natural Philosophy, University of Aberdeen

MS. received 1st *November* 1954

Abstract. The application of Stockbarger's technique to the growth of PbF_2 crystals is described. The melting point of PbF_2 was found to be 822°c ± 2°c, and owing to its fairly high vapour pressure at this temperature, the crystals were grown in an atmosphere of oxygen-free nitrogen at a pressure of 2 to 10 mm of mercury. The crystal structure was found to be of the fluorite type, and the lattice constant $5{\cdot}942 \pm 0{\cdot}001$ Å at 18°c. Refractive indices are given for various wavelengths in the visible spectrum. The transmission limits, defined as the wavelength at which a specimen of thickness 1 cm absorbs 50% of the incident radiation, are $11{\cdot}6\,\mu$ and 2800 Å.

§ 1. Introduction

Single crystals of lead fluoride (PbF_2) have been grown in this department in the past three years, and the infra-red transmission spectrum has previously been reported (Jones *et al.* 1952). Lead fluoride is useful as a prism material in visible and infra-red spectroscopy, and as a Cerenkov scintillator. The object of this paper is to describe the application of Stockbarger's method to the growing of lead fluoride crystals, in which special difficulties were encountered, and to give further information on the properties of the material.

§ 2. The Apparatus

The vacuum furnaces used for the growth of lead fluoride crystals were similar to the furnaces used for the growth of other fluorides, and they substantially follow those described previously by Stockbarger (1949). One furnace was capable of growing crystals up to 3·5 cm diameter, and the other up to 10 cm diameter.

The pressure inside the furnace chamber was measured by either a Pirani gauge ($1\,\mu$ to 1 mm Hg) or a Vacuostat gauge (0·5 mm to 10 mm). Temperatures were measured by chromel–alumel thermocouples situated in the upper and lower parts of the furnace. The pressure in the chamber was controlled by passing a stream of oxygen-free nitrogen into it via a needle valve.

The crucible was of the usual type, made from dense graphite, cylindrical in shape, with a conical base. It was provided with a close fitting graphite lid which had to be securely fitted to the crucible to avoid the risk of it blowing off and the powder escaping when the chamber pressure was reduced. A small hole (0·04 in. diameter) in the centre of the lid allowed the gases to escape from the crucible.

§ 3. The Furnace Atmosphere

The vapour pressure of lead fluoride at its melting point is about 1 mm Hg, so the pressure in the furnace chamber must be greater than this, otherwise the salt will sublime out of the crucible. The crystal has to be grown in an inert atmosphere of pressure of the order of a few millimetres of mercury or more.

Dry oxygen-free nitrogen was found to be suitable, and pressures in the range 2 to 10 mm Hg were satisfactory. Before entering the furnace chamber the gas had to be thoroughly dried by passing through a liquid air trap.

Unsatisfactory crystals, which were black and completely opaque, were produced when a static atmosphere of dry nitrogen was used. This was possibly due to the liberation of gases such as water vapour and carbon dioxide from the various components inside the furnace chamber, which accumulated in the furnace atmosphere. These gases reacted with hot lead fluoride and produced salts which were easily reduced to lead by the graphite. The lead stayed in the melt making the crystal formed black and opaque. It was necessary to maintain a continuous flow of nitrogen through the furnace during the whole crystallization process.

A static atmosphere of dry hydrogen fluoride gas was also tried, but again black crystals were produced. In this case also the gases liberated in the chamber would cause lead to form. Hydrogen formed by the action of the hydrogen fluoride on the metallic baffles would also reduce lead fluoride to lead.

In order to grow satisfactory crystals it was first of all necessary to dry thoroughly the raw material before heating was commenced. The graphite crucible was carefully cleaned and dried before it was filled with the salt. It was then placed in position in the furnace and evacuated for twenty-four hours at a pressure of 1 to 5 μ. The temperature was raised to about 70°c and held at this value for a further twenty-four hours. Heating was then continued to about 600°c at a pressure of 5 μ or less. The rate of heating was such that outgassing never caused the pressure to rise above about 5 μ. If the pressure was allowed to rise too much, the gas in the furnace attacked the metal baffles and shortened their life considerably. At 600°c nitrogen was admitted to the furnace at a rate sufficient to raise the pressure to between 50 μ and 100 μ. Near 700°c the diffusion pump was switched off and the nitrogen flow adjusted to give the required pressure. In order to limit the quantity of nitrogen used, the cocks in the pumping lines could be partially closed.

§ 4. TEMPERATURE OF FURNACE AND METHOD OF CONTROL

The melting point of lead fluoride was found to be 822°c ± 2°c and during growth the top half of the furnace was held at 870°c and the bottom at 770°c. These temperatures were measured 1 inch above and below the horizontal baffle which separated the two halves of the furnace. No elaborate temperature control mechanisms were required. It was sufficient to control the voltage of the supply to the heater. This was done by means of an automatic voltage stabilizer, giving a maximum output of 100 amp at 230 volts ± 1%. The large thermal inertia of the furnace enabled it in effect to average the output of the stabilizer over a long time, with the result that at about 800°c the furnace temperature remained steady within about 5°c. The stabilizer supplied a suitable variable transformer, the output of which was transformed down by a 230/40 volt autotransformer. The variable transformer was fitted with an automatic drive so that the furnace could be heated or cooled at a given rate.

When the temperatures were steady, the crucible was lowered into the bottom half of the furnace at the rate of 1 mm per hour until crystallization was complete. The furnace was then cooled steadily to room temperature over a period of three days.

Like the other fluorides, lead fluoride did not adhere to the graphite crucible, and the crystal was easily removed. Care had to be taken to ensure that the crystal had reached room temperature before admitting air to the furnace chamber, otherwise the crystal was liable to crack.

§ 5. CHEMICAL PURITY OF THE RAW MATERIAL

As in the growth of all crystals from the melt, chemical purity is important in the preparation of lead fluoride crystals. The metallic ion impurity should be low (probably not more than one part in 10^4), but it is the presence of electronegative ions such as the oxide, hydroxide, carbonate, nitrate, sulphate, acetate, etc., that has the more deleterious effect on the crystal. All these ions are converted to the oxide on heating. The oxide is then reduced by the graphite crucible to form lead which has a vapour pressure lower than that of lead fluoride at the melting point of the latter. The lead stays in the melt, rendering the crystal black and opaque throughout its normal transmitting range.

Attempts to grow crystals from ordinary commercial lead fluoride were unsuccessful, and 'scavenging' with ammonium fluoride did not improve the product although this technique worked well with lithium fluoride (Haven 1949, Jones 1951). This could be due to the fact that lead is comparatively resistive to the action of hydrogen fluoride. Mercuric fluoride was also tried as a 'scavenger' and it seemed to be a little more effective than ammonium fluoride, although the crystal was still far from satisfactory. Lead fluoride is itself used as a scavenger for other materials, and this ability to act as a scavenger probably makes it more difficult to grow as a pure substance.

British Drug Houses, Ltd., have manufactured recently an 'extra pure' grade which is suitable for growing crystals, and satisfactory crystals can be grown from this without the aid of 'scavengers'.

§ 6. CRYSTAL STRUCTURE OF LEAD FLUORIDE CRYSTALS GROWN FROM THE MELT

Kolderup (1924–25) discovered that lead fluoride was dimorphous and that there was a transition between the two modifications at 400°. The high temperature modification had the cubic fluorite structure. Schumann (1933) determined the edge of the elementary cell of the cubic phase to be $5{\cdot}942 \pm 0{\cdot}002$ Å, and also showed that the low temperature phase was orthorhombic. The density of the cubic modification was $7{\cdot}678$ g cm^{-3}, and the orthorhombic $8{\cdot}369$ g cm^{-3}. Ketelaar (1932) measured the cell dimensions of the latter modification to be $a = 3{\cdot}80 \pm 0{\cdot}01$ Å, $b = 6{\cdot}41 \pm 0{\cdot}02$ Å, $c = 7{\cdot}61 \pm 0{\cdot}03$ Å. Byström (1947) re-measured the lattice constants of both forms of lead fluoride. He prepared the low temperature form by the action of excess hydrogen fluoride on lead carbonate, and the cubic form by heating this to 400°c, and then cooling. His figures were $5{\cdot}935 \pm 0{\cdot}003$ Å for the cubic form, and $a = 3{\cdot}987 \pm 0{\cdot}002$ Å, $b = 6{\cdot}441 \pm 0{\cdot}003$ Å, $c = 7{\cdot}648 \pm 0{\cdot}004$ Å for the orthorhombic form.

The density of the crystals grown from the melt was determined to be $7{\cdot}763 \pm 0{\cdot}001$ g cm^{-3} at 18°c. This agrees fairly well with Schumann's value for the cubic form. The former value corresponds to a lattice constant of the cubic form of $5{\cdot}942$ Å, which is in perfect agreement with Schumann's value, and fair agreement with that obtained by Byström. X-ray powder photographs confirmed that lead fluoride crystallizes from the melt in the cubic fluorite structure, and stays in this form when cooled. The lattice constant calculated from these photographs was 5·942 ± 0·001 Å at 18°c. X-ray photographs also showed that the lead fluoride powder, as supplied by the manufacturers, was orthorhombic.

Lead fluoride crystals do not cleave as readily as the other fluorides with the fluorite structure into well-defined planes. Striking the crystal with a sharp blade never produces the cleavage planes characteristic of the other fluorides, and in this respect lead fluoride resembles a glass. Well-defined cleavage planes have been noticed, however, in crystals which have cracked due to thermal shock.

§ 7. REFRACTIVE INDICES OF LEAD FLUORIDE CRYSTALS

The refractive indices of lead fluoride crystals were determined using a prism of angle approximately 62·5°, and faces approximately 3 cm × 3 cm. Measurements were made on the Hilger–Watts research spectrometer. During the course of the measurements, the room temperature varied between 17·5°c and 19·8°c. A temperature correction was applied so that the values given are correct at 18·5°c. The correction was determined by varying the temperature of crystal between 17°c and 22°c, and observing the change in the angle of minimum deviation. In this temperature range the correction is given by

Line	Wavelength (Å)	μ	$d\mu/d\lambda$ (cm^{-1})
He red b	7065·20	1·75450	740
Hg red	6907·5	1·75577	795
He red	6678·15	1·75770	900
H_2 red C	6562·79	1·75873	960
Cd red	6438·47	1·75995	1015
Na D	5895·932	1·76626	1395
	5889·965	1·76635	
He Yellow d	5875·620	1·76653	
Hg Green e	5460·742	1·77290	1755
Cd Green	5085·822	1·78026	2240
He Blue	4921·928	1·78406	2505
H_2 Blue F	4861·327	1·78562	2620
Cd Blue	4799·908	1·78724	2740
Cd Violet	4678·151	1·79067	3010
He Violet	4471·482	1·79744	3590
H_2 Violet G	4340·465	1·80232	3940
Hg Violet h	4046·56	1·81554	

$$\mu_{T(^\circ C)}=\mu_{18\cdot5}\{1+0\cdot00005(18\cdot5-t)\}.$$

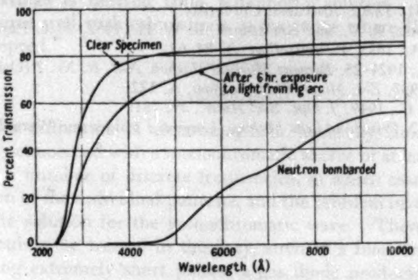
The refractive indices are believed to be correct to within 4 in the fifth decimal place. The values of $d\mu/d\lambda$ were read off from the graph of μ against λ. The reciprocal dispersive power, given by $(\mu_D-1)/(\mu_F-\mu_C)$, is 28·6 which is approximately the same as that of double extra dense flint glass.

§ 8. TRANSMISSION SPECTRUM

The long wave absorption edge of lead fluoride has previously been reported (Jones *et al.* 1952). The figure then given was that a specimen 1 cm thick absorbed 50% of the incident radiation at 11·1 μ. More recent specimens have shown a slightly better transparency, the 50% absorption limit being at 11·6 μ.

In the ultra-violet region the 50% absorption edge for a block 1 cm thick occurred at 2800 Å for the best crystals. Some slight variation in the position of the absorption edge was noticed in different specimens. The figure shows the shape of the edge for a crystal 1 cm thick. No correction has been applied for reflection losses.

Lead fluoride crystals become yellow-brown in colour after prolonged exposure to x-rays or ultra-violet rays. The colouring was due to increased absorption at the shorter wavelengths (see figure). No selective absorption bands were observed. One specimen was irradiated with neutrons in the pile at the Atomic Energy Research Establishment, and its spectrum is also shown in the figure. The transmission was very low up to 4500 Å, and there was a slight increase in absorption between 6500 and 7500 Å.



Transmission of PbF_2 in ultra-violet region.

§ 9. OPTICAL WORKING OF LEAD FLUORIDE CRYSTALS

Lead fluoride crystals have approximately the same hardness as lithium fluoride, and the same procedure for polishing was used (Twyman 1952). The crystals readily crack if subjected to even slight thermal shock, and it is essential that they should not be touched by hand during the grinding and the polishing. The crystal, the laps, the grinding and polishing materials, and particularly the water used, must all have reached the same temperature before an attempt is made to work the crystal.

The crystal is more easily worked if previously annealed. This is done by slowly heating the crystal to about 450°C over a period of three days, holding at this temperature for twenty-four hours and then cooling to room temperature over a period of three days. This must be carried out in a vacuum furnace because heating in the atmosphere ruins the optical clarity of the crystal.

§ 10. CONCLUSION

The fact that lead fluoride crystallizes from the melt in the cubic form, and also its high transparency and dispersion in the visible and near ultra-violet, makes it a very suitable prism material for spectroscopy in this region. It would also be suitable for use in the infra-red region up to about 11·5 μ. It is stable in the atmosphere and unaffected by moisture. The presence of the heavy lead atoms, and the fact that it is transparent to about 2800 Å, makes it suitable material for use in scintillator counters.

ACKNOWLEDGMENTS

I wish to express my thanks to Professor R. V. Jones for giving me the facilities to carry out this work, and for his help and encouragement during the many early failures to produce satisfactory crystals, to Dr. R. W. H. Stevenson for much useful discussion, and to Dr. T. Smith for help on the x-ray techniques. It is a pleasure to acknowledge the skill of our workshop staff, particularly Mr. T. Wratten who was mainly responsible for the construction of the apparatus.

REFERENCES

BYSTRÖM, A., 1947, *Ark. Kemi. Min. Geol.* A, **24,** 33.
HAVEN, Y., 1949, *Disc. Faraday Soc., Crystal Growth*, **5**.
JONES, D. A., 1951, *Thesis*, Aberdeen University.
JONES, D. A., JONES, R. V., and STEVENSON, R. W. H., 1952, *Proc. Phys. Soc.* B, **65,** 906.
KETELAAR, J. A. A., 1932, *Z. Kristallogr.* A, **84,** 62.
KOLDERUP, N. H., 1924–25, *Bergens Museum Harbok. Nat. R. Nr.* 2 (Oslo).
SCHUMANN, A., 1933, *Zbl. Miner. Geol. Paläont.* A, **122**.
STOCKBARGER, D. C., 1949, *J. Opt. Soc. Amer.*, **39,** 731.
TWYMAN, F., 1952, *Prism and Lens Making* (London : Hilger and Watts Ltd.), pp. 171–181.

The Diffraction of a Non-Monochromatic Electromagnetic Wave by a Slit and a Grating

By M. H. N. POTOK
Department of Electrical Engineering, The Royal Technical College, Glasgow

MS. received 25th June 1954, and in amended form 22nd October 1954

Abstract. A general solution for the diffraction of non-monochromatic waves by slits and gratings is derived from Kirchhoff's solution. The results of measurements agree well with calculations in the range where Kirchhoff's solutions can be accepted.

§ 1. Introduction

In dealing with diffraction of electromagnetic waves by slits and gratings, one is usually concerned with a monochromatic source or at least with sources radiating at a number of discrete frequencies, in which case the diffraction pattern is the sum of the individual patterns, and the problem involves a repeated application of the solution for the monochromatic wave. There are, however, sources which cannot be treated in this way, such as a monochromatic source which is radiating extremely short pulses, a gas diode producing noise over a wide band and a spark microwave generator (Potok 1953).

Such sources are often of interest, where a wide band is required either for its own sake or as a background from which a narrow band of any desired centre frequency can be filtered out. When dealing with these sources it may be necessary to disperse the spectrum by means of a slit or a grating. It is therefore often of interest to know what diffraction pattern will result from a given source.

§ 2. The Diffraction by a Slit

It is well known that the Kirchhoff solution of diffraction by apertures and obstacles is only approximately correct, the error becoming appreciable as the wavelength approaches the aperture dimensions. This is the state of affairs at microwave frequencies. On the other hand the Kirchhoff solution is relatively simple in comparison with more accurate solutions (Schwarzchild 1902, Morse and Rubenstein 1938, Stratton and Chu 1939, Bouwkamp 1954).

Figure 1 gives the relation between output power and diffraction angle for a slit width equal to one and two wavelengths λ, as calculated from Kirchhoff's solution and as found by measurements, performed by the author, using a monochromatic source. This shows that while there is some deviation when $a=\lambda$, the agreement is very close for $a=2\lambda$. Thus, as long as $a \geqslant 2\lambda$ one is reasonably justified in adhering to Kirchhoff's solution.

Consider an arbitrary wave $f(t)$ falling at right angles on a slit of width a as in figure 2. Let $F(t)$ be the diffracted wave at angle θ. Since by Kirchhoff's solution each element of the slit is a source of equal amplitude and phase, the wave

$$F(t)=\int_0^a f\left(t-\frac{x\sin\theta}{c}\right)dx$$

where x is measured along the width of the slit, and c is the velocity of propagation of light (Farrands and Brown 1954).

The power received at angle θ will now be

$$P=\int_{-\infty}^{\infty}|F(t)|^2\,dt.$$

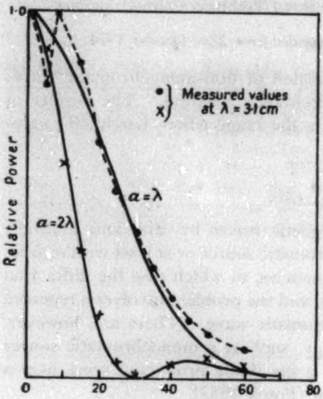


Figure 1.

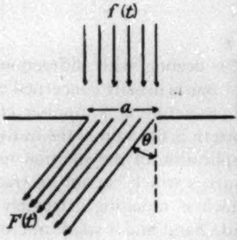


Figure 2.

Following the method described by Farrands and Brown we make use of Parseval's theorem

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}|F(t)|^2\,dt=\int_{0}^{\infty}|G(\omega)|^2\,d\omega$$

where $G(\omega)$ is the Fourier integral of $F(t)$.

Let

$$g_0(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}f(t)\exp(-j\omega t)\,dt;$$

then

$$G(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(t)\exp(-j\omega t)\,dt=\frac{1}{2\pi}\int_0^a\int_{-\infty}^{\infty}f\left(t-\frac{x\sin\theta}{c}\right)\exp(-j\omega t)\,dx\,dt$$

$$=\int_0^a\exp\left(-\frac{jx\sin\theta}{c}\right)dx\,\frac{1}{2\pi}\int_{-\infty}^{\infty}f(t)\exp(-j\omega t)\,dt$$

$$=\frac{c}{j\omega\sin\theta}g_0(\omega)\left\{\exp\left(-\frac{j\omega a\sin\theta}{c}\right)-1\right\}.$$

Writing $(a\sin\theta)/c=\eta$

then $$G(\omega)=\frac{a}{j\omega\eta}g_0(\omega)(1-\exp(-j\omega\eta));$$

hence $$|G(\omega)|^2=\frac{2a^2}{\omega^2\eta^2}|g_0(\omega)|^2(1-\cos\omega\eta)$$

and $$P=4\pi\int_0^{\infty}\frac{2a^2}{\omega^2\eta^2}|g_0(\omega)|^2(1-\cos\omega\eta)\,d\omega. \qquad \ldots\ldots(1)$$

As an example, consider a simple case of an exponentially decaying wave as radiated by a spark generator $f(t)=\exp(-\alpha t)\sin\omega_0 t$.

This case has been dealt with by Bartenstein (1909) but led him to a very complex solution.

It can be shown that

$$|g_0(\omega)|^2=\frac{\omega_0{}^2}{4\pi^2(\omega-\omega_0-j\alpha)(\omega-\omega_0+j\alpha)(\omega+\omega_0-j\alpha)(\omega+\omega_0+j\alpha)}$$

hence $$P=\frac{2a^2\omega_0{}^2}{\pi\eta^2}\int_0^\infty\frac{(1-\cos\omega\eta)\,d\omega}{\omega^2(\omega-\omega_0-j\alpha)(\omega-\omega_0+j\alpha)(\omega+\omega_0-j\alpha)(\omega+\omega_0+j\alpha)} \qquad \ldots\ldots(2)$$

which can be solved by Cauchy contour integration over a semicircle of infinite radius indented at origin, leading to:

$$\begin{aligned}P=\frac{a^2\omega_0{}^2}{2\pi\eta^2}\Bigg[&\frac{\eta\pi}{2(\omega_0{}^2+\alpha^2)}+\frac{\pi}{8\omega_0\alpha(\omega_0-j\alpha)^3}+\frac{\pi}{8\omega_0\alpha(\omega_0+j\alpha)^3}\\&-\frac{\pi\exp\{j(\omega_0+j\alpha)\}}{8\omega_0\alpha(\omega_0+j\alpha)^3}-\frac{\pi\exp\{-j(\omega_0-j\alpha)\}}{8\omega_0\alpha(\omega_0-j\alpha)^3}\Bigg]\\=\frac{\omega_0c^2}{2\alpha(\omega_0{}^2+\alpha^2)\sin^2\theta}\Bigg[&\omega_0{}^3-3\omega_0\alpha+\frac{2a\sin\theta\,\omega_0\alpha(\omega_0{}^2+\alpha^2)}{c}\\&-(\omega_0{}^2+\alpha^2)^{3/2}\exp\left(-\frac{\alpha a\sin\theta}{c}\right)\cos\left(\frac{\omega a\sin\theta}{c}-\phi\right)\Bigg]\end{aligned}$$

where $$\phi=\tan^{-1}\frac{\alpha}{\omega_0}\,\frac{3\omega_0{}^2-\alpha^2}{\omega_0{}^2-3\alpha^2}.$$

Writing $\alpha=f_0\delta$ (δ is the logarithmic decrement of the function $f(t)$), and $\omega_0=2\pi f_0$,

$$\begin{aligned}P=\frac{c^2}{8\pi^2f_0{}^3\delta(1+\delta^2/4\pi^2)^3\sin^2\theta}\Bigg[&1-\frac{3\delta^2}{4\pi^2}+\frac{2a\delta(1+\delta^2/4\pi^2)^{1/2}\sin\theta}{\lambda}\\&-(1+\delta^2/4\pi^2)^{3/2}\exp\left(-\frac{a\delta\sin\theta}{\lambda}\right)\cos\left(\frac{2\pi a\sin\theta}{\lambda}-\phi\right)\Bigg].\end{aligned}$$

When $\theta=0$, the expression (2) simplifies to

$$P_0=4a^2\pi\int_0^\infty|g_0(\omega)|^2\,d\omega=\frac{a^2\omega_0{}^2}{4\alpha(\omega_0{}^2+\alpha^2)}=\frac{a^2}{4\delta f_0(1+\delta^2/4\pi^2)}.$$

Setting $P_0=1$ and substituting into P gives

$$\begin{aligned}P=\frac{\lambda^2}{2a^2\pi^2(1+\delta^2/4\pi^2)^2\sin^2\theta}\Bigg[&1-\frac{3\delta^2}{4\pi^2}+\frac{2a\delta(1+\delta^2/4\pi^2)\sin\theta}{\lambda}\\&-\left(1+\frac{\delta^2}{4\pi^2}\right)^{3/2}\exp\left(-\frac{a\delta\sin\theta}{\lambda}\right)\cos\left(\frac{2\pi a\sin\theta}{\lambda}-\phi\right)\Bigg].\end{aligned} \qquad \ldots\ldots(3)$$

As a check, if $\delta=0$, then

$$P_{\delta=0}=\frac{\lambda^2}{2a^2\pi^2\sin^2\theta}\left(1-\cos\frac{2a\pi\sin\theta}{\lambda}\right)=\frac{\sin^2\{(a\pi\sin\theta)/\lambda\}}{(a^2\pi^2\sin^2\theta)/\lambda^2}$$

which is the usual formula quoted in textbooks.

It has been assumed in the above that the response of the detector is flat over the whole bandwidth of the radiation. If that is not the case, then the expression (1) has to be multiplied inside the integral by a function giving the frequency response of the detector. Figure 3 gives the plot of relative power against angle of diffraction for the logarithmic decrement $\delta = 0$, 0·25, 0·5, 1 when $a = 2\lambda$, $\lambda = 3{\cdot}1$ cm as well as the results of a measurement on radiation from a spark generator whose characteristic by other measurements was found to be $\lambda_0 = 3{\cdot}4$ cm, $\delta = 0{\cdot}27 \pm 0{\cdot}03$.

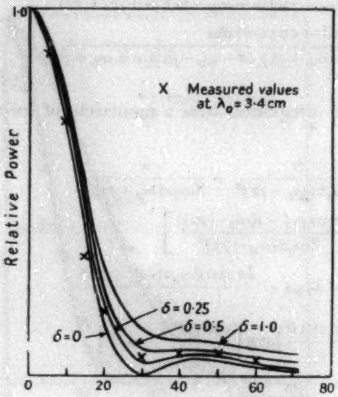


Figure 3.

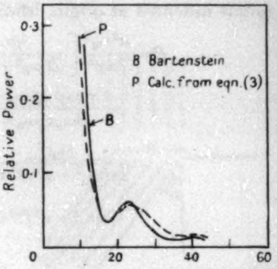


Figure 4.

As a further illustration figure 4 reproduces the output obtained by Bartenstein (1909) with $\lambda = 8{\cdot}66$ cm, $a = 30$ cm as well as the curve calculated on the assumption that $\delta = 0{\cdot}25$ with the help of expression (3).

§ 3. DIFFRACTION BY A GRATING

Consider here gratings in which the reflecting surfaces have the same width as the slits and the distance between the centre lines of two slits (grating constant) is $2a$. The available equipment produced a beam which could be taken as uniform at the most over three slits which set the limit to the experiment.

Figures 5 and 6 compare the diffraction by two and three slits respectively between measurements with 3·1 cm monochromatic waves and Kirchhoff's solution. Applying the method described in §2 to calculate the diffraction pattern by a grating when the incident wave is $f(t)$, one can write the diffracted wave as

$$F(t) = \int_0^a f\left(t - \frac{x \sin\theta}{c}\right) dx + \int_{2a}^{3a} f\left(t - \frac{x \sin\theta}{c}\right) dx + \ldots .$$

Hence, if $g_0(\omega) = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{\infty} f(t) \exp(-j\omega t)\, dt$

$$G(\omega) = \frac{a}{j\omega\eta} g_0(\omega)[1 - \exp(-j\omega\eta) + \exp(-2j\omega\eta) - \ldots$$
$$+ \exp\{-(2k-1)j\omega\eta\}]$$

where $\eta = a \sin\theta / c$ and k = number of slits.

This leads to
$$|G(\omega)|^2 = |g_0(\omega)|^2 \frac{c^2}{\omega^2 \sin^2\theta} \frac{1-\cos 2k\omega\eta}{1+\cos\omega\eta}.$$

This expression is not readily integrable and it is preferable to leave it in the form of a series, in which case
$$|G(\omega)|^2 = \frac{a^2}{\omega^2\eta^2} |g_0(\omega)|^2 [2k + \sum_{n=1}^{2k-1} 2n(-1)^n \cos(2k-n)\omega\eta];$$

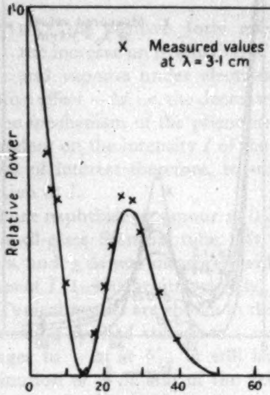


Figure 5.

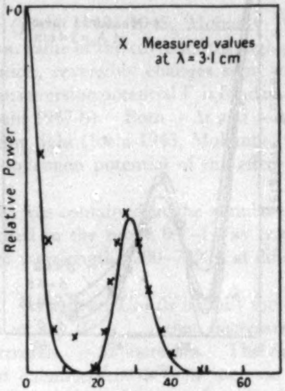


Figure 6.

hence by the same method as described earlier
$$P_g = \frac{8\pi a^2}{\omega^2\eta^2} \left\{ \int_0^\infty k |g_0(\omega)|^2 \, d\omega + \sum_{n=1}^{2k-1} n(-1)^n \int_0^\infty \exp[j(2k-n)\omega\eta] |g_0(\omega)|^2 \, d\omega \right\}. \qquad \ldots\ldots(4)$$

Again, as an example, consider the case when $f(t) = \exp(-\alpha t) \sin \omega_0 t$. Then
$$P_g = \frac{c^2}{8\pi^2 f_0^3 \delta (1+\delta^2/4\pi^2)^3 \sin^2\theta} \left\{ k\left(1 - \frac{3\delta^2}{4\pi^2}\right) \frac{2ka\delta(1+\delta^2/4\pi^2)^{1/2} \sin\theta}{\lambda} \right.$$
$$\left. + (1+\delta^2/4\pi^2)^{3/2} \sum_{n=1}^{2k-1} n(-1)^n \exp\left(-\frac{(2k-n)a\delta\sin\theta}{\lambda}\right) \cos\left[(2k-n)\frac{2\pi a \sin\theta}{\lambda} - \phi\right] \right\}$$

where
$$\phi = \tan^{-1}\left(\frac{\delta}{2\pi} \frac{3-\delta^2/4\pi^2}{1-3\delta^2/4\pi^2}\right).$$

Setting $P_g = 1$ at $\theta = 0$, gives
$$P_g = \frac{\lambda^2}{2a^2k^2\pi^2(1+\delta^2/4\pi^2)^2 \sin^2\theta} \left\{ k\left(1 - \frac{3\delta^2}{4\pi^2}\right) + \frac{2ka\delta(1+\delta^2/4\pi^2)^{1/2}\sin\theta}{\lambda} \right.$$
$$+ (1+\delta^2/4\pi^2)^{3/2} \sum_{n=1}^{2k-1} n(-1)^n \exp\left[\frac{(2k-n)a\delta\sin\theta}{\lambda}\right]$$
$$\left. \times \cos\left[(2k-n)\frac{2\pi a\sin\theta}{\lambda} - \phi\right] \right\}. \qquad \ldots\ldots(5)$$

This expression allows one to compute the solution for any number of slits. This has been shown for $a=\lambda$ and $\delta=0$, 0·25, 0·5 and 1 for two and three slits in figures 7 and 8 which also include the results obtained with a spark generated wave, the same as in § 2.

It is interesting to note that the maximum of the first order does not occur at the angle $\theta'=\sin^{-1}\lambda/2a$ as would be the case with a large number of slits.

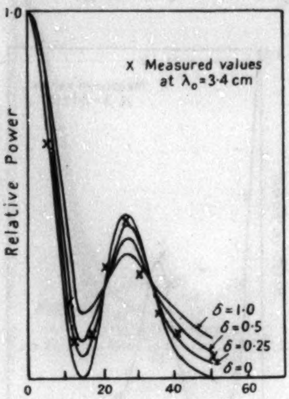


Figure 7.

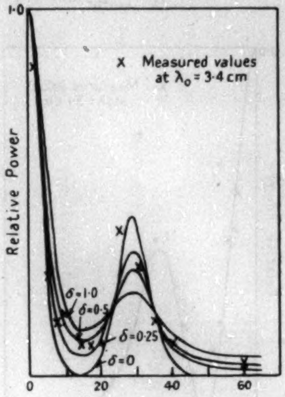


Figure 8.

§ 4. Conclusions

The paper derives a simple solution of the problem of diffraction of non-monochromatic waves by slits (expression (1)) and by gratings (expression (4)). The correctness of the solution depends on the degree to which Kirchhoff's approach to diffraction can be accepted.

References

Bartenstein, D., 1909, *Ann. Phys., Lpz.* (Series 4), **29,** 301.
Bouwkamp, C. J., 1954, *Rep. Progr. Phys.* (London : Physical Society), **17,** 35.
Farrands, J. L., and Brown, J., 1954, *Wireless Engr,* **31,** 81.
Morse, P. M., and Rubenstein, P. J., 1938, *Phys. Rev.,* **54,** 895.
Potok, M. H. N., 1953, *Brit. Instn Radio Engrs,* **13,** 490.
Schwarzchild, K., 1902, *Math. Ann.,* **55,** 177.
Stratton, J. A., and Chu, L. J., 1939, *Phys. Rev.,* **56,** 99.

RESEARCH NOTES

Inversion Potential of the Joshi Effect and Light Intensity

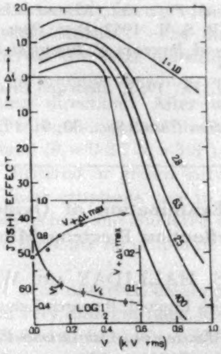
By S. R. MOHANTY and P. SRINIVASA RAO
Physico-Chemical Laboratories, Banaras Hindu University, India

MS. received 21st September 1954 and in amended form 9th December 1954

That the positive Joshi effect $+\Delta i$ (Joshi 1943, 1945, Mohanty 1953), the increase on irradiation of the r.m.s. value of the current i through gases and vapours under electrical excitation, reversibly changes sign, to the negative effect $-\Delta i$, i.e. the decrease of i, at an inversion potential V_i is fundamental for the mechanism of the phenomenon (Joshi 1947 b). Both $+\Delta i$ and $-\Delta i$ are dependent on the intensity I of the irradiating light (Joshi 1943, Mohanty 1954). It was of interest therefore, to study the inversion potential of the effect as a function of I.

Pure naphthalene vapour at 0·2 mm Hg was contained in the annular space of an all-glass Siemens tube. It was excited in the range 0·1–1·0 kv (r.m.s.), 50 c/s, and $\pm\Delta i$ was measured with light of wavelength 3700–7800 Å at different values of I (1–400, arbitrary units).

Typical results are shown in the figure. At a given I, $+\Delta i$ initially increases with rising applied voltage to a maximum at, say, $V_{+\Delta i\,\max}$, then decreases and changes to $-\Delta i$ at V_i; at still larger potentials, $-\Delta i$ increases. The rate of diminution of $+\Delta i$ and of the subsequent enhancement of $-\Delta i$ increase with increasing I. The range of potentials in which Δi is positive, on the other hand, is inversely proportional to I. Furthermore, whilst $V_{+\Delta i\,\max}$ (inset to figure) is increased, V_i diminishes.



Inversion potential of the Joshi effect at different light intensities.

According to Harries and von Engel (1951), the external light is strongly absorbed by, and causes dissociation of, the molecules, e.g. chlorine, in an adsorbed layer on the electrode walls. Attachment by the atoms of the secondary electrons from these walls causes $-\Delta i$. The secondary electron emission is caused by positive ions and photons, the γ and $\eta\theta g$ processes (Loeb 1939, Meek and Craggs 1953). Whilst electron attachment would diminish the extent to which these processes occur, their enhancement is independent of the external light. The observed $+\Delta i$, and the simultaneous occurrence of $+\Delta i$ and $-\Delta i$ (Joshi 1947 b, Jatar 1950, Khastgir and Setty 1952), cannot therefore be accounted for by the above mechanism. The theory of the effect due to Joshi (1946, 1947 a, 1947 b), on the other hand, postulates electron liberation, due to the external light, from an activated layer of ions and neutral molecules on the interior walls covered by the external electrodes. Capture of these photoelectrons by the electronegative elements in the discharge forms slow moving negative ions which reduce i mainly through accumulation in space. The photoelectrons which escape capture, and their secondaries, provided the field is favourable, cause $+\Delta i$. Along the ascending part of a curve (see figure), the effect is purely positive. The potential $V_{+\Delta i\,\max}$ marks the onset of the simultaneous occurrence of $+\Delta i$ and $-\Delta i$. The observed resultant $+\Delta i$ therefore decreases. At V_i, $+\Delta i=-\Delta i$, and consequently $\pm\Delta i=0$. With increase in I, $-\Delta i$, occurring alone or simultaneously with $+\Delta i$, increases (Mohanty 1954). For inversion to obtain at a larger I, $+\Delta i$ must increase proportionately. This is achieved by a diminution in V toward $V_{+\Delta i\,\max}$, and therefore in V_i.

ACKNOWLEDGMENT

Grateful thanks of the authors are due to Professor S. S. Joshi for his kind interest in the work.

REFERENCES

HARRIES, W. L., and VON ENGEL, A., 1951, *Proc. Phys. Soc.* B, **64,** 916.

JATAR, D. P., 1950, *J. Sci. Industr. Res.* B, **9,** 283.

JOSHI, S. S., 1943, *Proc. Indian Sci. Congr., Chem. Sec.*, Presidential Address; 1945, *Proc. Indian Acad. Sci.* A, **22,** 389; 1946, *Proc. Indian Sci. Congr., Phys. Sec.*, Abstract 26; 1947 a, *Ibid., Phys. Sec.*, Abstract 25; 1947 b, *Curr. Sci.*, **16,** 19.

KHASTGIR, S. R., and SETTY, P. S. V., 1952, *Proc. Phys. Soc.* B, **65,** 823.

LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York: John Wiley).

MEEK, J. M., and CRAGGS, J. D., 1953, *Electrical Breakdown of Gases* (Oxford: The Clarendon Press).

MOHANTY, S. R., 1953, *J. Indian Chem. Soc.*, **30,** 9; 1954, *J. Phys. Chem.*, **58,** 178.

The Examination of Oxide Films by Reflection Electron Microscopy

BY J. S. HALLIDAY AND W. HIRST

Associated Electrical Industries Limited, Research Laboratory, Aldermaston, Berkshire

MS. received 13*th October* 1954

THIS note describes an investigation by reflection electron microscopy of the change in surface topography of metals when oxidized. The Metropolitan-Vickers EM3 electron microscope modified as described by Haine and Hirst (1953) was used. The problem was suggested by an investigation of the frictional behaviour of oxide films on metals by Hirst and Lancaster (1954) who showed that for a given metal, e.g. copper, the protection afforded by an oxide film could depend on the rate at which the oxide film grows. Optical examinations of the films revealed no differences which might account for their frictional behaviour and films having greatly different frictional characteristics were shown by electron diffraction to be crystallographically alike. Reflection electron microscopy, however, shows that topographical differences exist which satisfactorily account for the different mechanical properties.

Figure 1 (Plate) shows the effect of oxidation on the surface structure of aluminium. Figure 1(*a*) is a reflection micrograph of an electropolished surface, the viewing angle θ_2 is 3·9° and the angle of illumination θ_1 is approximately 0·1°. As may be seen, the general surface is flatter than 0·1° although there are slight undulations which cause intensity variations from place to place; these occur most particularly at the grain boundaries. There are only a few surface features which are smaller in size than the crystal grains. Figure 1(*b*), $\theta_1=0{\cdot}1°$, $\theta_2=3{\cdot}9°$, is a micrograph obtained from a similarly electropolished specimen which had been oxidized in air at 500°c for 30 minutes; this treatment produced a thin film of amorphous Al_2O_3 which was not detectable by eye or under the optical microscope. From figure 1(*b*) it can be seen that the general surface, over areas comparable with the grain size, remains flat to better than 0·1° but the film has acquired a somewhat granular appearance. Nevertheless, because of the extremely small illuminating angle, the granular appearance does not denote the presence of very appreciable surface slopes, but only of slopes just exceeding 0·1°. In fact the smaller hillocks are of about 0·5 micron diameter and only 60 Å high; there are in addition a few larger surface features but their height does not exceed 0·05 micron. It appears that this film adheres well to the underlying metal surface and very largely maintains the initial surface topography.

In contrast, a different type of growth behaviour occurred when oxidizing nickel, as shown in figure 2 (Plate). Figure 2(*a*) shows a nickel surface which had been abraded on several successively finer grades of emery paper and finally polished on 'Selvyt' cloth with 'Brasso'. The micrograph was taken with the electron beam directly across the final polishing direction and $\theta_1=0{\cdot}2°$, $\theta_2=3{\cdot}8°$. There are traces of the previous abrading treatment and also the surface undulates across the polishing direction. After oxidizing a similar nickel specimen in air at 500°c for 15 minutes a yellow film of NiO formed; figure 2(*b*) was obtained with this specimen ($\theta_1=0{\cdot}2°$, $\theta_2=3{\cdot}8°$). It will be seen that the oxide film has either flaked, blistered or grown outwards in solid mounds on many parts of the surface. These eruptions vary in size, the largest appearing in figure 2(*b*) is 5 microns across its base and about 0·5 micron high. The general appearance of the oxide where there are no eruptions is, however, similar to the oxidized aluminium.

Both types of growth have been observed on copper. Two films of similar thickness, which were grown at very different rates, are shown in figure 3 (Plate). Figure 3(*a*) shows an electropolished surface after oxidation at 300°c for 10 minutes ($\theta_1=0{\cdot}2°$, $\theta_2=3{\cdot}8°$). According to Miley (1937) this treatment produces a Cu_2O film approximately 4000 Å thick. The surface film is very similar in appearance to the oxidized nickel surface shown in figure 2(*b*). The eruptions were visible even when θ_1 was raised to 4°. Figure 3(*b*) ($\theta_1=0{\cdot}4°$, $\theta_2=3{\cdot}6°$) shows the surface of a finely abraded copper specimen which had been oxidizing in air at room temperature for eighteen months. The Cu_2O film thickness was about 3000 Å. There are surface abrasion marks running across the direction of the electron beam but there are no eruptions visible. Using electron diffraction, a randomly oriented Cu_2O pattern was obtained from both sorts of surface film, the only slight difference being that the contrast of the diffraction rings against the diffuse background was lower from the low temperature oxide. This lower contrast might indicate the presence of amorphous oxides but could also be due to the general flatness of the oxide.

There is some evidence that the eruptions in figures 2(*b*) and 3(*a*) were blisters rather than solid mounds. Whilst some copper and mild steel specimens, having thick oxide films on them, were being examined in the electron microscope, many of the larger eruptions appeared to bubble and split. After splitting, the loose flakes curled away and finally became stationary. These phenomena could be explained if the oxide film had blistered on the metal surface thus leaving cavities between the oxide and parent metal. Due to the heating effect of the electron beam and because of the thermal isolation and small thermal capacity of the oxide, its temperature would rise more rapidly than that of the underlying metal. Thus excessive compressive stresses would arise in an oxide blister which may finally cause it to burst. The presence of blisters in the oxide films correlates with indirect evidence from measurements of the oxidation rates of various metals, for example Pilling and Bedworth (1923) found abrupt increases at irregular intervals in the oxidation rate of copper in air at 500°c. These abrupt increases were assumed to arise from a blistering and cracking of the brittle oxide film thus causing virgin metal to be exposed to the atmosphere.

There is the possibility that some features on the heat-treated surface might be due to thermal etching effects. However, the possibility is very remote since the heating was done in air at atmospheric pressures and the temperatures were somewhat low. Also, of course, any thermal etching would be extremely hindered by the presence of relatively thick oxide films.

As mentioned earlier this work was prompted by experiments of Hirst and Lancaster (1954) who found that the frictional force arising when sliding a clean copper hemisphere, 1 cm diameter, on a flat finely abraded copper surface oxidized at room temperature for nine months, corresponded to a coefficient of friction $\mu = 0{\cdot}3$, even for loads as high as 10 kg. A similar experiment using a similarly prepared flat but oxidized at 300°c for 10 minutes, and having a thicker oxide film than the former, gave a value $\mu > 1{\cdot}0$ which is about the same as that obtained using a freshly abraded flat. There is, therefore, as is already recognized no clear relationship between the protective action of an oxide film and its thickness. Finch (1950) has stated that the crystallographic character of the film has also to be considered; the oxides formed on many metals at room temperature are amorphous whereas those formed at higher temperatures are often crystalline. Finch (1950) suggests that amorphous films are in general more protective than crystalline ones. Whilst this may be true, the two oxide films used by Hirst and Lancaster were crystallographically similar and it seems more likely that the difference in the mechanical behaviour arises from some other cause. The most probable cause is the topographical difference revealed by reflection electron microscopy. Thus, the tenacious oxide formed at room temperature will deform with the underlying copper, but the high temperature oxide will fracture whenever the slider strikes a blister so that the slider will move on to a clean metal surface or one which is covered by such a thin oxide film as to be unable to prevent rupture and intermetallic welding. This view is supported by the fact that the film on aluminium oxidized at higher temperatures (500°c for 30 minutes) was exceptional in that it retained the surface topography of the underlying metal and exceptional also in that it alone, of the metals examined, gave better protection against frictional damage than did the thinner oxide film forming at room temperature.

ACKNOWLEDGMENTS

The authors wish to thank Mr. M. E. Haine for the use of the electron microscope and Dr. T. E. Allibone for permission to publish this note.

REFERENCES

FINCH, G. I., 1950, *Proc. Phys. Soc.* A, **63,** 785.
HAINE, M. E., and HIRST, W., 1953, *Brit. J. Appl. Phys.*, **4,** 239.
HIRST, W., and LANCASTER, J. K., 1954, *Proc. Roy. Soc.* A, **223,** 324.
MILEY, H. A., 1937, *J. Amer. Chem. Soc.*, **59,** 2626.
PILLING, N. B., and BEDWORTH, R. E., 1923, *J. Inst. Met.*, **29,** 529.

The Magnetic Susceptibility of Metallic Neodymium

BY L. F. BATES, S. J. LEACH, R. G. LOASBY AND K. W. H. STEVENS
Department of Physics, The University, Nottingham

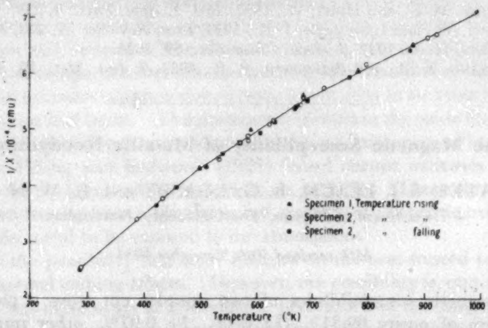
MS. received 30th November 1954

THE magnetic susceptibilities of two samples cut from a piece of neodymium of purity 99·81% (Li 0·1%, Fe 0·07%, other impurities less than 0·01%) have been measured in the temperature range 290°K to 1000°K. The material was kindly supplied by Dr. A. R. Powell of Johnson Matthey and Co. Ltd. The apparatus used was that described by Bates and Hughes (1954). The specimen was 2 cm long and of square cross section 0·2 × 0·2 cm, mounted in a sheath of tantalum foil of thickness 0·002 cm whose magnetic effect was neglected. The temperature was measured with a calibrated platinum–platinum-10% rhodium thermocouple mounted on the outside of the tantalum sheath to avoid contamination. The rate of change of the temperature of the specimen during the experiments was always less than 50° per hour. In the region 500°K to 700°K the rate of change was less than 10° per hour, and one hour was allowed for stabilization at a stated temperature before making any individual measurement.

The absolute mass susceptibility was measured with the apparatus described by Bates and Mallard (1950), but using an aqueous solution of nickel chloride as a standard (Nettleton and Sugden 1939), and the value found at 21·5°c was $38{\cdot}6 \pm 0{\cdot}4 \times 10^{-6}$, giving an atomic susceptibility of 5560×10^{-6} e.m.u. To test for ferromagnetic impurity the susceptibility was measured, at room temperature, in fields up to 4500 oersteds. No variation of susceptibility was found greater than the experimental error (±1%).

The variation of susceptibility with temperature is given in the figure which shows the reciprocal mass susceptibility plotted against temperature.

From 290°K to 500°K the graph can be well represented by the equation $\chi = C/(T-\theta)$, where $\theta = -15$°K and $C = 1{\cdot}724$, giving an atomic moment equal to $3{\cdot}72\mu_B$ in close agreement with the theoretical value for a trivalent neodymium ion $3{\cdot}68\,\mu_B$ (Van Vleck 1932). This is also in agreement with the results of Elliot, Legvold and Spedding (1954) who, over the temperature range 145°K to 297°K, found $\theta = -16$°K and an atomic moment of $3{\cdot}68\,\mu_B$. Above 500°K this representation no longer holds, and since there do not appear to be any changes in crystal structure below 1000°K (Spedding and Daane 1954) some explanation in terms of changes in population of levels seems likely. In the trivalent ions the first excited level, $4f^3\,{}^4I_{11/2}$, lies about 1800 cm^{-1} above the ground level, $4f^3\,{}^4I_{9/2}$. If we assume that in the metal a similar description of the 4f electrons is valid, the susceptibility can be closely fitted by assuming the gap is reduced to approximately 1350 cm^{-1}.



Temperature variation of susceptibility of metallic neodymium.

References

Bates, L. F., and Hughes, D., 1954, *Proc. Phys. Soc.* B, **67,** 28.
Bates, L. F., and Mallard, J. R., 1950, *Proc. Phys. Soc.* B, **63,** 520.
Elliot, J. F., Legvold, S., and Spedding, F. H., 1954, *Phys. Rev.*, **94,** 50.
Nettleton, H. R., and Sugden, S., 1939, *Proc. Roy. Soc.* A, **173,** 313.
Spedding, F. H., and Daane, A. H., 1954, *Journal of Metals* (May), 504.
Van Vleck, J. H., 1932, *The Theory of Electric and Magnetic Susceptibilities* (Oxford: University Press).

Breakdown Effect in p–n Alloy Germanium Junctions

By R. D. KNOTT, I. D. COLSON and M. R. P. YOUNG
Research Laboratories of The General Electric Company, Limited, Wembley, England

MS. received 5th *November* 1954

Early measurements of McAfee *et al.* (1951) on the breakdown effect in p–n junctions under high reverse voltage conditions showed that the slope of the current–voltage characteristic in the voltage-limiting region was in essential agreement with that predicted from an extension of Zener's theory of internal field emission (1934). Furthermore, injection of carriers into the very narrow space charge region of the grown germanium junctions used in this work produced a unit quantum yield. The absence of charge multiplication is implicit in the field emission theory. Under the above conditions, the experimental value quoted for the maximum potential gradient in the junction at breakdown was 2×10^5 v cm^{-1}, as compared with the theoretically predicted value of $6{\cdot}9 \times 10^5$ v cm^{-1}.

More recent work of McKay and McAfee (1953) has shown that charge multiplication factors as high as eighteen may occur in the pre-breakdown region of grown junctions having a greater barrier width, the multiplication factor increasing rapidly as the voltage-limiting region is approached. Carriers injected by photons, alpha-particles or thermal generation are multiplied in the same manner, and approximately equal multiplication factors apply for injected electrons and for injected holes. The experimental data were well represented by ionization rates computed by conventional avalanche theory, or the Townsend β process, of gas discharge physics (Loeb 1939). For such a model, it was shown that, with equal rates of ionization of electrons and holes, and a uniform field, we have

$$\alpha_0 = \frac{1}{w_{\mathrm{eff}}}(1 - 1/M), \qquad \ldots\ldots(1)$$

where α_0 is the rate of ionization, i.e. the number of electron–hole pairs produced by an incident carrier per centimetre path traversed in the direction of the field, M is the observed multiplication factor, and w_{eff} the 'effective' barrier width for the process.

To deduce α_0 as a function of the field, the effective barrier width was arbitrarily selected as one-half the actual barrier width and the effective 'uniform' field was taken to be the maximum in the space-charge region.

Two forms of breakdown mechanism were thus recognized, an avalanche breakdown in broad junctions, and internal field emission in junctions which are too narrow to support appreciable multiplication. The above processes will now be examined in connection with indium bead type, germanium p–n alloy junctions. For this case, we may consider an abrupt planar transition from p- to n-type conductivity located at the alloying interface. Furthermore, the charge density in the p-type material is much greater than in the original n-type region, a saturation resistivity of 0·003 Ω cm for the recrystallized layer being quoted (Pankove 1954), as compared with a practical range of n-type resistivity of 0·2–40 Ω cm. Under these conditions, the potential in the space-charge region is of the parabolic type discussed by Schottky, and application of Shockley's theory of the p–n junction (1949) gives the relationships

$$C/A = 1{\cdot}33 \times 10^4/(\rho_n \psi)^{1/2}\ \mathrm{pF\ cm^{-2}} \qquad \ldots\ldots(2)$$

$$w = 1{\cdot}41/(C/A)\ \mathrm{cm}, \quad E_{\max} = 2\psi/w\ \mathrm{v\ cm^{-1}} \qquad \ldots\ldots(3)$$

where C/A is the barrier region capacity per unit area, w the barrier width, $E_{\max}$ the maximum field in the barrier region, ρ_n the resistivity of the n-type material (Ω cm), and ψ the applied potential (v), the dielectric constant of germanium being taken as 16 and the electron mobility in n-type material as 3600 cm sec^{-1} per v cm^{-1}.

Figure 1 contains, firstly a family of curves relating ψ and C/A for several values of ρ_n, according to equation (2), secondly the curve A relating the reverse voltage for field emission breakdown to C/A, assuming the breakdown field of 2×10^5 v cm quoted by McAfee *et al.* (1951) and, finally, the breakdown curve B according to avalanche theory which was derived by extrapolation of the data relating α_0 and $E_{\max}$ published by McKay and McAfee (1953). In this case we have defined the 'breakdown' condition as $M \to \infty$, i.e. from equation (1) $\alpha_0 w_{\text{eff}} = 1$, and retained the arbitrary definitions of 'effective' barrier width and 'uniform' field mentioned above. The two curves intersect at $\rho_n \simeq 0{\cdot}5\ \Omega$ cm so that for resistivities above this value, avalanche breakdown would be predicted and, for lower resistivities, the breakdown should be controlled by the field emission effect.

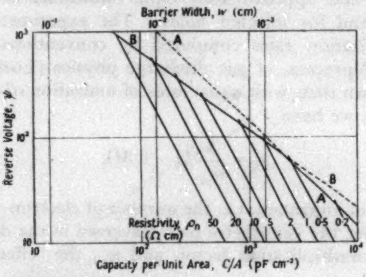


Figure 1. Predicted breakdown voltage of germanium p–n alloy junctions. A : breakdown due to internal field emission. B : breakdown due to carrier multiplication in the barrier region.

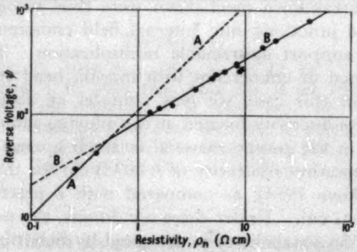


Figure 2. Variation of germanium p–n alloy junction breakdown with resistivity. A and B : predicted from field emission and carrier multiplication theory, respectively. Points represent the maximum breakdown voltage observed for a given resistivity.

In the present work, slices of single crystal n-type germanium, with known resistivities in the range 0·2–40 Ω cm, were cut perpendicular to the 110 direction and eight junctions, approximately 1 mm in diameter, were made on each slice using the well established alloying technique with indium as the p-type impurity. The breakdown potentials were determined by applying an adjustable 50 c/s a.c.

voltage across the junctions and suitable series resistors and displaying the reverse current–voltage characteristics on a calibrated cathode ray oscilloscope. A variation of some 20% in the value of the breakdown voltage was found over a batch of normal junctions made on base material of a given resistivity and, in all cases, the maximum observed value was accepted as that most probably associated with the quoted resistivity.

The results are plotted in figure 2, the breakdown curves A and B of figure 1 being redrawn as functions of the practical variables ρ_n and ψ. Curve A, based on the field emission theory, gives a breakdown potential $\psi_B = 112{\cdot}5\rho_n$, $\rho_n < 0{\cdot}5\,\Omega$ cm, and curve B for the case of avalanche breakdown $\psi_B = 83{\cdot}4\rho_n^{\,0{\cdot}61}$, $\rho_n > 0{\cdot}5\,\Omega$ cm. Further points in the resistivity range $0{\cdot}4$–$1{\cdot}0\,\Omega$ cm would be required to provide a detailed knowledge of the form of the curve in the transition region, but the good agreement between the predicted and observed values indicates that the above theory may be equally well applied to alloy type as to grown germanium p–n junctions.

REFERENCES

LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York: John Wiley).

MCAFEE, K. B., RYDER, E. J., SHOCKLEY, W., and SPARKS, M., 1951, *Phys. Rev.*, **83**, 650.

MCKAY, K. G., and MCAFEE, K. B., 1953, *Phys. Rev.*, **91**, 523.

PANKOVE, L., 1954, *R.C.A. Rev.*, **15**, 75.

SHOCKLEY, W., 1949, *Bell. Syst. Tech. J.*, **28**, 435.

ZENER, C., 1934, *Proc. Roy. Soc.* A, **145**, 523.

LETTERS TO THE EDITOR

The Computation of Electric Dipole Moments

Le Fèvre (1953) in the last edition of *Dipole Moments* writes " Hence dielectric constants and densities should normally be measured with the same precision ". I believe that there is a real danger of this view being misunderstood. It is hoped that the following remarks will help to clarify the situation.

It is convenient to begin by defining an experimental quantity Q by

$$Q \equiv \left\{\frac{d}{dc}\left(\frac{\epsilon-1}{\epsilon+2}\right)-\frac{d}{dc}\left(\frac{n^2-1}{n^2+2}\right)\right\}_{c\to 0} \qquad \ldots\ldots(1)$$

where ϵ denotes dielectric constant, n refractive index and c moles per volume. It has been pointed out by Smith (1950) that the same quantity Q can be determined by several alternative formulae, in particular

$$Q=\frac{3}{(\epsilon_1+2)^2}\left(\frac{d\epsilon}{dc}\right)_{c\to 0}-\frac{3}{(n_1{}^2+2)^2}\left(\frac{dn^2}{dc}\right)_{c\to 0} \qquad \ldots\ldots(2)$$

where the subscript 1 relates to the pure solvent, or

$$Qd_1=\frac{3}{(\epsilon_1+2)^2}\left(\frac{d\epsilon}{dm}\right)_{m\to 0}-\frac{3}{(n_1{}^2+2)^2}\left(\frac{dn^2}{dm}\right)_{m\to 0} \qquad \ldots\ldots(3)$$

where m denotes moles per mass of solvent and d_1 the density of the pure solvent, or

$$Q\frac{d_1}{M_2}=\frac{3}{(\epsilon_1+2)^2}\left(\frac{d\epsilon}{dw}\right)_{w\to 0}-\frac{3}{(n_1{}^2+2)^2}\left(\frac{dn^2}{dw}\right)_{w\to 0} \qquad \ldots\ldots(4)$$

where w denotes mass-fraction of solute and M_2 molar mass of solute. Formulae (1), (2), (3) and (4) are physically and mathematically equivalent. The accuracy with which Q can be determined is comparable with that with which $\epsilon-\epsilon_1$ can be measured over the range of composition in which ϵ is approximately linear in the concentration. This accuracy is rarely closer than 1%. It is desirable to determine c, or d_1 and m, or d_1 and w with a precision of one part in a thousand but higher accuracy is pointless.

So much for the experimental determination of Q; we must now consider its theoretical interpretation. It has been shown (Guggenheim 1949) that Debye's fundamental formulae are physically and mathematically equivalent to the relation.

$$Q=\frac{4\pi \boldsymbol{N}}{3}\left(\frac{\mu^2}{3\boldsymbol{k}T}+\gamma_a-\gamma_a{}'\right) \qquad \ldots\ldots(5)$$

where μ denotes the dipole moment of the solute molecule, γ_a denotes the atomic contribution to the polarizability of the solute molecule and $\gamma_a{}'$ is defined by

$$\frac{4\pi \boldsymbol{N}}{3}\gamma_a{}'=\left(\frac{\epsilon_1-1}{\epsilon_1+2}-\frac{n_1{}^2-1}{n_1{}^2+2}\right)V_2 \qquad \ldots\ldots(6)$$

where V_2 denotes the partial molal volume of the solute. If with Le Fèvre we write

$$ {}_0P_2 = \frac{4\pi N}{3}\frac{\mu^2}{3kT} \qquad \ldots\ldots(7)$$

then

$$ {}_0P_2 = Q - \frac{4\pi N}{3}(\gamma_a - \gamma_a') \qquad \ldots\ldots(8)$$

Both γ_a and γ_a' will usually be small compared with $\mu^2/3kT$. Usually little is known about γ_a and it is evidently pointless to be more meticulous about γ_a' than about γ_a. Consequently quite a rough estimate of V_2 is sufficient, such as its value in the pure solute.

Accurate density measurements lead to a more exact value of V_2, which we have seen is uncalled for, and they lead to nothing else!

In the absence of information concerning γ_a it is largely a matter of taste what to do about it. Le Fèvre's recommendation is to neglect γ_a so that (8) reduces to

$$ {}_0P_2 = Q + \frac{4\pi N}{3}\gamma_a' . \qquad \ldots\ldots(9)$$

My suggestion (Guggenheim 1949) is that since $\gamma_a - \gamma_a'$ is likely to be smaller than γ_a it is simpler and no less reasonable to neglect $\gamma_a - \gamma'_a$ so that (8) reduces to

$$ {}_0P_2 = Q . \qquad \ldots\ldots(10)$$

For many solvents, e.g. benzene and carbon tetrachloride, the difference between ϵ_1 and n_1^2 is sufficiently small to allow the approximation

$$Q \simeq \frac{3}{(\epsilon_1+2)^2}\left\{\left(\frac{d\epsilon}{dw}\right)_{w\to 0} - \left(\frac{dn^2}{dw}\right)_{w\to 0}\right\}$$

or other related approximations. Since, however, the saving in calculation due to such approximations is trivial and since their introduction has caused serious confusion between *these approximations* and our *other exact relations*, they are better ignored and I regret ever having suggested their use. In any case such approximations should not be used when the solvent is dioxane. An example of the correct application of our formulae to measurements in dioxane is given elsewhere (Guggenheim and Prue 1954).

University of Reading, E. A. GUGGENHEIM.
20th October 1954.

GUGGENHEIM, E. A., 1949, *Trans. Faraday Soc.*, **45**, 714.
GUGGENHEIM, E. A., and PRUE, J. E., 1954, *Physico-chemical Calculations* (Amsterdam : North-Holland Publishing Company), problem 38, p. 106.
LE FÈVRE, R. J. W., 1953, *Dipole Moments*, 3rd Edn (London : Methuen), p. 59.
SMITH, J. W., 1950, *Trans. Faraday Soc.*, **46**, 394.

REVIEWS OF BOOKS

The Sun, edited by GERARD P. KUIPER. Pp. xix + 745. (Chicago: University Press; London: Cambridge University Press, 1953.) £4 14*s*.

" But if we wish to know what a star really is we must approach it closely, and this is possible only in the case of the sun " (Hale 1915): the monumental work now published gives ample justification for the belief that both then and now the study of the sun is one of the surest means of progress in Astrophysics. Under the editorship of Professor G. P. Kuiper (Chicago) twenty-three workers in six different countries have combined to give " a systematic and comprehensive account of our present knowledge concerning the sun . . ."; this is Volume I of a series of four designed to cover the whole solar system. The book contains nine chapters; the first eight, each with one exception the work of single authors, cover observational results and theory, while the ninth consists of 13 short articles on different observational techniques. It is a general reference work written at a research level and the presentation is designed for anyone " versed in the physical sciences . . . .". On the whole, the needs of the specialist have been put first, over 1000 references are given, most of the main articles are long and contemporary work is often given in its entirety. The non-specialist would probably be grateful in several instances for stricter selection and more emphasis.

Solar physics as a whole with some history of the subject is surveyed in an introduction by L. Goldberg (Michigan). The description and interpretation of the visible features of the solar surface and the solar spectrum (in the optical range), occupying half the book, are contained in 'The Photosphere' by M. Minnaert (Utrecht), a short chapter on 'The identification of solar lines' by C. E. Moore (Bureau of Standards), 'The Chromosphere and the Corona' by H. C. van de Hulst (Leiden) and 'Solar Activity' by K. O. Kiepenheuer (Freiburg). Professor Minnaert deals with the continuous spectrum, the wavelength distribution of energy and the variation across the disc, he compares the empirical distribution of pressure and temperature in the solar atmosphere thereby obtained with theoretical photospheric models. In a section on the Fraunhofer lines he considers the form of the atomic absorption coefficient in the photosphere and the line profiles resulting from the transfer of radiation through the solar atmosphere. Professor Minnaert, who with his collaborators has produced an invaluable 'photometric solar atlas', is aware, as few others, of the possibilities and pitfalls of spectrophotometry; the emphasis in this article, however, is on the theoretical interpretation of results and not on observational problems and techniques. Professor van de Hulst has contributed a lively and comprehensive article; he draws many different types of observation together in a unified picture of the outer layers of the sun. His frequent critical surveys and conclusions at the ends of sections are a great help to the reader. Professor Kiepenheuer deals with sunspots, faculae, chromospheric flares, prominences, corpuscular emission, the correlation of these phenomena with terrestrial events, and the solar component of cosmic rays. The subject requires the presentation of a wide range of observational material, and Dr. Kiepenheuer succeeds in giving a clear and convincing idea of the present position of our knowledge.

'The Sun as a Star' by B. Strömgren (Yerkes and McDonald Observatories) deals with the physics of the unobservable solar interior, the energy production by thermonuclear processes under solar conditions, and the deduction of physical properties at all radial distances. J. L. Pawsey and S. F. Smerd (Sydney) contribute a chapter on 'Solar Radio Emission' and bring out well the possibilities and present limitations of this second 'window' in the solar spectrum. Chapter 8 on 'Solar Electrodynamics' by T. G. Cowling (Leeds) is somewhat different in character from the earlier ones, being a discussion of "the hydrodynamics of solar material in the presence of a magnetic field. Under this general title are discussed solar rotation, the origin of solar magnetic fields, sunspots and the solar cycle, and electromagnetic theories of flares". Professor Cowling's approach to this new subject is more frankly didactic than that of the earlier writers; the reviewer, definitely 'non-specialist' in the subject, can only admire a stimulating and extremely interesting article.

Techniques and observational problems are deferred until the final chapter which is the work of 13 authors in a little over 100 pages. To an observational astronomer the emphasis and method of treatment are somewhat disappointing. Lyot's coronagraph and the recent work on solar magnetic fields by H. W. and H. D. Babcock stand out with the distinction they deserve. But for the rest, articles by experts on specialized equipment, though excellent in themselves, do not give the reader a clear picture of the achievements and problems on the observational side, nor do they enable him to assess the experimental reliability of the results on which the earlier chapters are based. While a single author could not have had such specialized knowledge, a much more balanced picture would have resulted; as it is, the fundamental work on the solar constant, for instance, is allotted only 5 pages, the design of solar telescopes very little more, and practical solar spectroscopy (optical range) only 3 pages.

The subject of solar physics is just about 100 years old, but in this volume attention is understandably concentrated on the last 25 years and especially on the post-war period. One could wish, not only for sentimental reasons, that a slightly larger proportion of space had been allowed for earlier work which laid the foundations for the present outlook. Such criticisms as may be made are, however, minor; for this comprehensive and authoritative statement of the present position physicists, solar and terrestrial, will be extremely grateful. They will certainly need *The Sun* available for reference, and the solar physicists will undoubtedly wish to possess copies of their own.

M. G. ADAM.

Faster than Thought, edited by B. V. BOWDEN. Pp. xix+416. (London: Pitman, 1953.) 35*s*.

The subject of automatic digital calculating machines has three aspects, the first concerned with the engineering aspects of design, construction and maintenance, the second with the means of using them to carry out significant calculations, and the third with those applications of such machines which have been made or are contemplated. This volume, written by a group of contributors under Dr. B. V. Bowden, who himself contributes seven substantial chapters, as editor, gives a survey of the whole subject covering all three aspects, though the treatment of the second is slight compared with that of the first and third.

About half the contributors are associated with Messrs. Ferranti Ltd., or with the Manchester University Computation Laboratory where there is a Ferranti machine installed. This leads to a slight but not serious lack of balance. One result is that terminology which is appropriate to the Manchester machine but which is peculiar to the Manchester group (such as ' line ' for a storage location or register) is used without comment as if it were generally accepted, and some engineering features are emphasized at the expense of others (for example, there is little mention of the use of magnetic materials for storage other than in the form of a ' magnetic drum '). These, however, are details; and when allowance is made for them, the book is an admirable survey for the general reader, with emphasis on the British contribution to the subject.

Much of the book is written in a lively and stimulating style, and this applies particularly to Dr. Bowden's own contributions, which cover a wide range and are written with a light touch and a mastery of the frivolous-sounding example to make a serious point; they will be a fruitful source of quotation for writers or speakers on the subject. He includes a fuller account than has appeared elsewhere of Lady Lovelace and her interest and remarkable understanding of Babbage's projected ' Analytical Engine ' and a discussion of the processes used by experts in mental arithmetic in addition to the chapters more directly relating to recent developments in hardware and applications. An interesting and valuable appendix is a full reprint of Lady Lovelace's translation of Menebrea's 1842 article on Babbage's machines, with her notes (more than twice as long as the article itself). The book ends with a glossary which is both useful and entertaining, though it might well have been fuller.

In a work with a number of contributors, it is difficult to make sure that the contributions are fully cordinated or that every point which should be mentioned is included. In the work under review, the reader unfamiliar with finite differences may be confused by the reference on p. 8 to a table of differences on p. 142, as the ' differences ' in the table on p. 142 are not finite differences in the sense of p. 8 (they appear to be the results of a check of some kind, but their relation to the other numbers tabulated is not explained). Also there is no explanation of what is meant by an ' address ' in this context, and the term does not appear in either the glossary or the index, although it is used in the text in its technical sense; although the reader may well work out for himself the significance of the term, it seems important enough to deserve an explanation in the text and inclusion in the glossary; it is surely more important than the musical exploits of the Manchester machine (which, incidentally, are mentioned in the chapter on ' Organization of a typical machine ': is this facility really typical ?). Such lapses, however, are few, and the book as a whole is eminently readable, and read with care it will both stimulate interest and provide thought.

D. R. HARTREE.

Relaxation Methods, by D. N. DE G. ALLEN. Pp. ix + 257. (New York, Toronto, London: McGraw-Hill, 1945.) 53*s*. 6*d*.

It is almost twenty years since Southwell introduced the method of ' systematic relaxation of constraints ' for stress calculations in frameworks. This method was essentially one for solution of a set of simultaneous algebraic equations by successive approximation, though expressed in a way appropriate to the structural engineering context in which it was first presented. The method has subsequently been applied by Southwell and his colleagues, and by others, to problems in contexts widely different from the original context, in particular to the finite-difference forms of ordinary and partial differential equations; but some of the original terminology has remained.

Mr. Allen is one of those who have been most closely connected both with extending the range of the relaxation method, and with spreading practical knowledge of the use of it through the medium of summer schools in the subject. In this book he has written, out of this experience, a valuable practical guide to anyone having calculations to do which can be aided by the use of this technique.

The first two chapters are concerned with linear algebraic equations, treated without giving a physical interpretation to the quantities occurring in them, so that the reader is introduced to the essential ideas without the distraction of a physical context with which he may not be familiar. This is now, I consider, the right approach to a subject which is essentially one of general numerical technique rather than a part of any one field of pure or applied science. The third chapter is concerned with the application of frameworks in connection with which the method was originally devised. Most of the remaining thirteen chapters are concerned with applications to differential equations of various kinds, and end with a short chapter on a procedure for carrying out a relaxation calculation in three dimensions without having recourse to a pack of sheets of paper. Most of the differential equations, ordinary and partial, considered are linear; a non-linear equation is considered on p. 100, but it is not referred to as such (and it seems to have been overlooked that it can be made linear by a change of variable); there is no entry ' non-linear equations ' in the index.

A surprising feature is the absence of any mention in the text of Fox's method of improving the solutions of differential equations, obtained by relaxation, by including higher differences in the replacement of derivatives by finite differences, the only method mentioned of decreasing the truncation errors of finite difference formulae is the use of smaller intervals. The relevant paper of Fox is given in the bibliography, but without at least a reference to it, the reader whom it might help would be unlikely to find it. Fox's methods deserve at least a section, if not a chapter, and it is to be hoped that this omission will be rectified in a future edition.

However, apart from this gap, the book can be recommended as a guide to the practice of the use of relaxation methods.

D. R. HARTREE.

CONTENTS OF SECTION A

PAGE

Dr. W. D. Corner and Dr. G. H. Hunt. The Temperature Dependence of Magnetostriction in a Nickel Crystal 133

Dr. R. B. Potts. Combinatorial Solution of the Triangular Ising Lattice . . 145

Dr. A. C. Hurley. On the Method of Atoms in Molecules 149

Mr. R. M. Curr. The Coulomb Scattering of High-Energy Electrons and Positrons by Nuclei 156

Mr. S. J. Biel and Dr. E. H. S. Burhop. The Effect of Finite Nuclear Size on Bremsstrahlung Production 165

Prof. D. R. Bates and Mr. J. T. Lewis. Inelastic Heavy Particle Collisions Involving the Crossing of Potential Energy Curves: III—Charge Transfer from Negative Ions of Atomic Hydrogen to Protons 173

Mr. W. R. McMurray and Mr. C. H. Collie. The Radium Equivalent of ^{24}Na Sources and the Photodisintegration Cross Section of Deuterium. . 181

Dr. A. M. Lane. Studies in Intermediate Coupling: III—The Lithium Isotopes 189

Dr. A. M. Lane. Studies in Intermediate Coupling: IV—Further Remarks on the Nuclei ^{13}C and ^{13}N 197

Dr. P. B. Treacy. The Spin, Energy and Lifetime of ^{8}Be 204

Prof. Kazuhisa Tomita. On the State of Solid Hydrogen 214

Mr. R. Berg and Dr. L. Wilets. On the Validity of the Weizsäcker Inhomogeneity Correction Term 229

Mr. D. W. Sida. The Detachment of Electrons from Negative Hydrogen Ions by impact with Neutral Atoms 240

Dr. I. P. Grant. Theory of (d, p) and (d, n) Reactions: II—Coulomb Corrections and Numerical Results 244

Letters to the Editor:

Mr. J. M. Baker and Dr. B. Bleaney. The Nuclear Spin and Magnetic Moment of Terbium 159 257

Dr. M. Tinkham. Paramagnetic Resonance Absorption by the Ferrous Ion 258

Dr. K. M. Gatha and Mr. G. Z. Shah. The Total Cross Sections for the Nuclear Scattering of High Energy Nucleons 259

Dr. P. C. Vaidya. A New Static Spherically Symmetric Solution of Einstein's Unified Field Theory 260

Reviews of Books 263

Contents of Section B 264

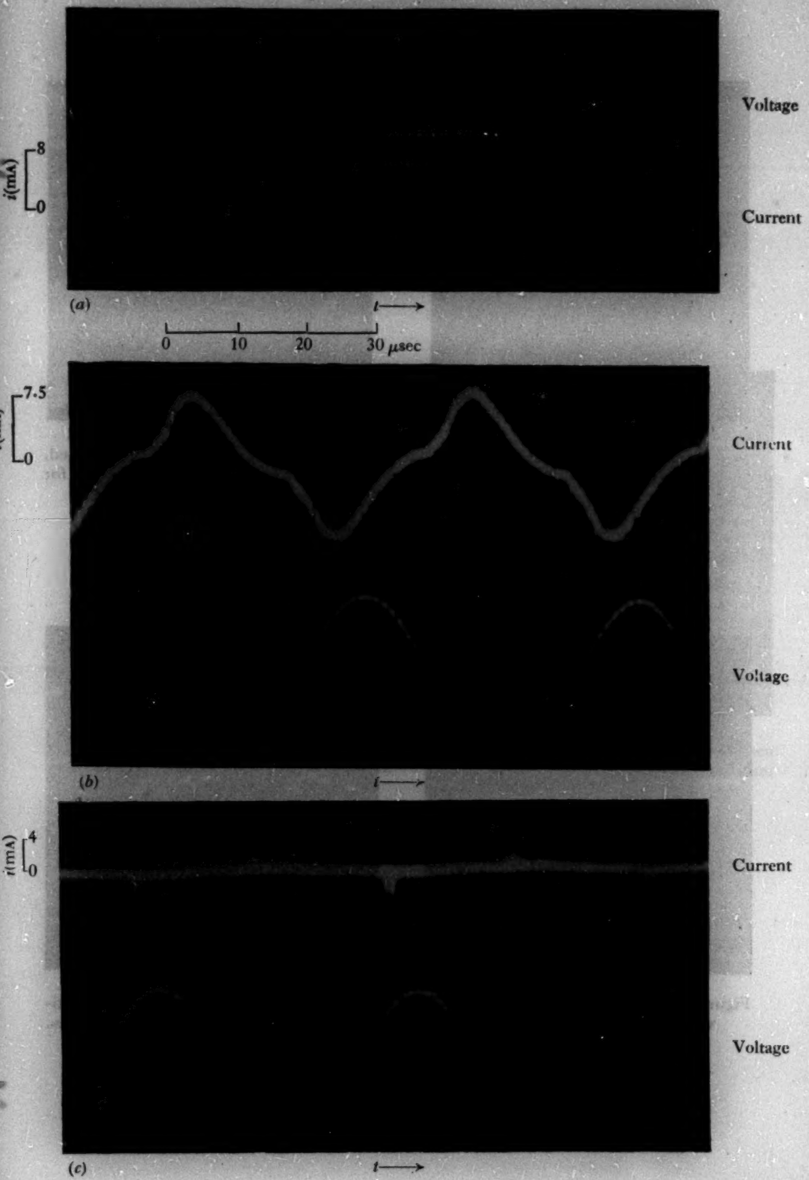


Figure 5. All these oscillograms were taken with a double beam oscilloscope, so that the current trace appears inverted with respect to the voltage trace. Thus a positive voltage, but a negative current, appear as upward deflections. Length of Pyrex vessel 1·8 cm. (*a*) $\lambda=67$ m. 60% excess potential applied (λ cut-off$=72$ m), $p=3$ mm Hg. (*b*) $\lambda=1400$ m, $p=27$ mm Hg. (*c*) $\lambda=7500$ m, and similar for all longer wavelengths, $p=27$ mm Hg. Peak voltage (*a*) 800 v, (*b*) 2280 v, (*c*) 2200 v.

Figure 1 (*a*). Aluminium, freshly electropolished. $\theta_1 \sim 0\cdot 1^\circ$, $\theta_2 \sim 3\cdot 9^\circ$.

Figure 1 (*b*). Aluminium, electropolished, then oxidized in air at 500°c for 10 minutes. $\theta_1 \sim 0\cdot 1^\circ$, $\theta_2 \sim 3\cdot 9^\circ$.

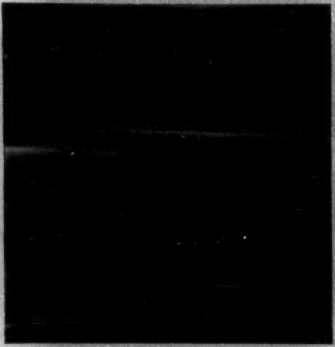


Figure 2 (*a*). Nickel, polished on 'Selvyt' with 'Brasso'. $\theta_1 = 0\cdot 2^\circ$, $\theta_2 = 3\cdot 8^\circ$.

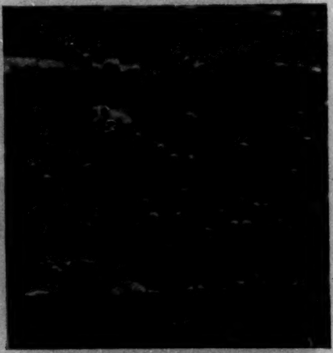


Figure 2 (*b*). Nickel, polished, then oxidized in air at 500°c for 15 minutes. $\theta_1 = 0\cdot 2^\circ$, $\theta_2 = 3\cdot 8^\circ$.

The scales on all these figures are equal.

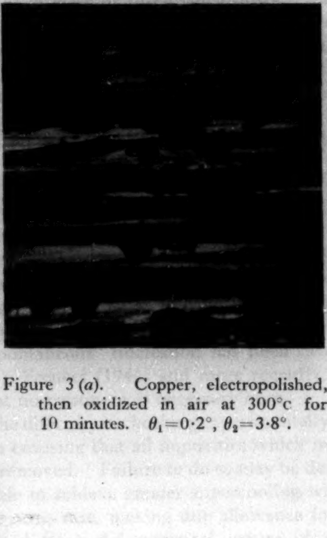


Figure 3 (*a*). Copper, electropolished, then oxidized in air at 300°C for 10 minutes. $\theta_1=0{\cdot}2^\circ$, $\theta_2=3{\cdot}8^\circ$.

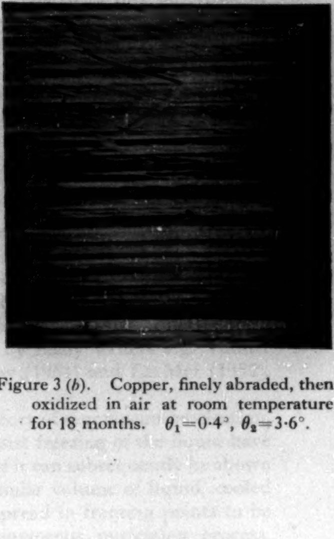


Figure 3 (*b*). Copper, finely abraded, then oxidized in air at room temperature for 18 months. $\theta_1=0{\cdot}4^\circ$, $\theta_2=3{\cdot}6^\circ$.

The Freezing of Supercooled Water

By S. C. MOSSOP†
Clarendon Laboratory, Oxford

Communicated by G. M. B. Dobson ; MS. received 9th September 1954
and in amended form 3rd December 1954

Abstract. To investigate the freezing of supercooled water of high purity, measurements were made of the freezing point of samples of water in glass capillary tubes, both with and without a hydrophobic coating on the walls. These showed the difficulty of excluding freezing nuclei that originate in room air. A simple technique was therefore devised for preparing uncontaminated specimens of water in glass or fused silica tubes. Samples prepared in this way froze at temperatures consistently lower than any previously recorded for similar volumes of water. The lowest value obtained for water in glass was −34·5°c for a volume of 6×10^{-4} cm^3. It is shown that some of the results of previous workers can be accounted for by the presence of foreign freezing nuclei in the water used. The relationship between volume and freezing point for spontaneous nucleation is discussed in the light of experimental results.

§ 1. Introduction

It is generally held that the freezing of a supercooled liquid is initiated by suitable impurities known as 'freezing nuclei', or in their absence by the chance formation of one or more aggregates of molecules which are large enough to be stable as nuclei of the new solid phase. These then grow until the liquid is completely frozen. The theory of the latter process of 'homogeneous' or 'spontaneous' nucleation has been described by many writers e.g. Volmer (1939), Frenkel (1946) and more recently, Bradley (1951) and La Mer (1952), so that no treatment is necessary here.

The difficulty in checking experimentally any theory of spontaneous nucleation, lies in ensuring that all impurities which might assist freezing of the liquid have been removed. Failure to do so may be detected if it can subsequently be shown possible to achieve greater supercooling with a similar volume of liquid, cooled at the same rate, making due allowance for the spread in freezing points to be expected from the statistical nature of the spontaneous nucleation process. It is thus of great interest for our knowledge of the mechanics of freezing, to find the lowest temperature to which liquids can be supercooled, and to find how this varies with volume and cooling rate.

§ 2. Historical Survey

The lowest reliable measurements of the freezing point of water so far recorded are those of Cwilong (1947) and Fournier d'Albe (1949) who obtained a value of approximately −41°c for the onset of freezing among water drops of about 1 μ diameter in cloud chambers. A similar threshold value (−38·9°c) was found by Schaefer (1946) for rather larger drops (diameter about 10 μ).

---

† Now at the South African National Physical Laboratory, Pretoria.

It is desirable to extend these observations and to find the lowest temperature to which water can be supercooled, over a wider range of volumes than can be obtained in cloud chambers. Numerous investigators have tackled this problem and a short account of relevant results will be given.

Meyer and Pfaff (1935) found that by filtering water through collodion filters or by successive distillation, samples could be produced which froze at −33°c. The volumes were stated to be small, though values are not given.

Heverly (1949) and Dorsch and Hacker (1950) found that the freezing points of drops resting on solid surfaces decreased with decreasing diameter. Their water samples were not, apparently, free from impurities, as later workers have recorded greater supercooling for the same droplet sizes.

Much lower temperatures were reported by Lafargue (1950) for 1 to 20μ droplets suspended on spiders' threads in silicone oil, or resting upon a mercury surface in silicone. Drops of distilled water and of various salt solutions froze at $-40{\cdot}5 \pm 1{\cdot}5$°c.

Hosler and Hosler (1952) made a series of investigations on water in Pyrex tubes of diameters ranging from 0·25 mm to 3 mm. Ordinary laboratory distilled water was used, and the containing tubes were cleaned by prolonged washing with chromic acid, followed by distilled water. They found that the freezing points of their samples were independent of the length, volume or surface area of the samples and depended only upon the diameter of the containing tube. The relationship between drop diameter and freezing point is shown in figure 1.

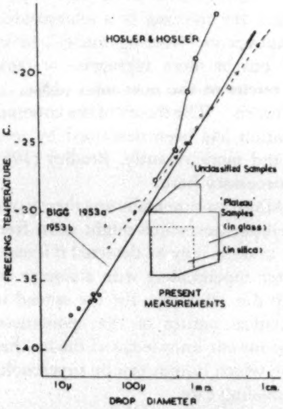


Figure 1. Relationship between drop diameter and freezing temperature of supercooled water. Present measurements on water in capillaries are plotted according to 'equivalent diameter', i.e. the diameter of a spherical drop of equal volume.

Because of the lack of any relationship between the volume of water and its freezing point, these authors discount the possibility of nucleation by the chance presence of suitable freezing nuclei. For the same reason they reject the Volmer theory of spontaneous nucleation. Instead they invoke a hypothesis due to Weyl (1951) that the formation of an electrical double layer at the surface of a water drop will render spontaneous nucleation more difficult owing to "distortion of the normal atomic structure of water". The smaller the radius, the greater will this effect be, and thus the greater the possible supercooling. The only geometrical parameter affecting the freezing temperature will thus be the maximum distance it is possible for any point in the liquid to be from the nearest surface.

The next investigation to which attention must be drawn was that of Wylie (1953), who determined the freezing point of water condensed into comparatively large glass tubes, elaborate precautions being taken to ensure that the glass was clean and that all impurities from the air of the room were removed. The freezing points of specimens, some as large as 6 cm^3 in volume, varied from $-7\cdot9$ to $-30 \pm 1^\circ\text{c}$. Wylie thought that nucleation was caused by minute crystals formed on the glass surface as a result of the heating and cleaning processes to which it had been subjected.

The most recent work was that of Bigg (1953 a), who studied the freezing of drops in the diameter range 0·1 mm to 2 cm. The drops of distilled water were suspended at the interface of two insoluble liquids, one heavier and one lighter than water. Arranging the drops into groups according to diameter, and plotting mean freezing point of a group of drops against mean diameter, he obtained the curve designated 'Bigg 1953 a' in figure 1. Results consistent with these were obtained with drops resting on a hydrophobic film of silicone oil baked on a glass surface. In a later paper (Bigg 1953 b) some results for very small drops, down to $20\,\mu$ diameter, were reported and as a result the curve was shifted to 'Bigg 1953 b'.

The temperature distribution of the freezing points of approximately 1000 droplets of 1 mm diameter was also analysed. Bigg showed that this agreed closely with a probability distribution, the parameters of which he deduced from the curve of freezing point against diameter. He concluded that the freezing of a drop was a probability event, depending upon the volume and the rate of cooling. A decrease by a factor of 10 in the rate of cooling from the usual value of 0·5 deg min^{-1} was calculated to cause a rise of approximately 2°c in the freezing point. This was in fact verified by cooling 164 1-mm drops at a rate of 0·05 deg min^{-1}.

In his later paper Bigg claims that since the same mean freezing temperature was obtained with distilled water from various sources, the impurities present must have been without effect upon the freezing point, i.e. freezing was due to spontaneous nucleation. Applying his results to cloud droplets he shows that freezing nuclei may be unimportant in glaciation in clouds below a temperature of -20°c; their effect will be masked by the spontaneous freezing of the cloud droplets. Where large drops are present, as in cumulus, this limit may be even higher.

Since this result is of great importance in our understanding of the processes of natural rain formation and also in possible stimulation of rain artificially, it is important that it should be checked by another method. The present research was therefore initiated in an attempt to investigate spontaneous nucleation phenomena using small quantities of carefully distilled water in sealed glass capillaries.

§ 3. Experiments on Freezing of Water in Glass Capillaries with Hydrophobic Coating

In an attempt to avoid any possible influence the walls might have on the freezing point, they were coated with a hydrophobic layer. The relationship between droplet volume and freezing point was investigated using capillaries of approximately constant diameter and varying the length of the thread of water in them.

3.1. *Preparation of Capillaries*

The capillary tubes were freshly drawn from Pyrex to an internal diameter of 0·20 to 0·30 mm. Immediately after being drawn the tubes were treated in order to cover them internally with a hydrophobic layer. Three different coatings were used, namely, dimethyl dichlorosilane, 'Experimental Silicone D 3032' and 'Drisil 29' (Preparations of Midland Silicones Ltd).

3.2. *Distillation of Water*

The water used in the experiments was obtained by redistillation of ordinary laboratory distilled water in an all-Pyrex still. Steam was first passed through the still for about 15 minutes before any distillate was collected.

3.3. *Filling of Tubes*

The distillate reservoir had a side tube covered by a thin rubber diaphragm. The capillaries could be passed through a tiny hole in this diaphragm directly into the distillate, so that ingress of room air was prevented as far as possible. Freshly distilled water was first sucked through each tube for about 5 minutes so that it had been flushed by many times its own volume of water. Water was then drawn up into the tube in such a way that, in addition to the test drops, there would be a short thread at each end to prevent contamination of the test drop from room air before sealing. The capillary was then sealed by applying an oxy-gas flame between the blocking and test drops.

To obtain a valid comparison between the freezing points of drops of different volumes, each capillary had two threads of water drawn into it, one about 6 cm long and the other about 2 mm, spaced a few millimetres apart.

3.4. *Determination of Freezing Points*

The freezing points were found by placing the specimens in a Dewar vessel containing acetone, which was then cooled by a hoop of copper rod having one end dipping in the acetone and the other in an adjacent Dewar of liquid oxygen. In this way the cooling rate could be kept very close to the desired value of 0·5 deg min^{-1} which was used throughout these experiments unless otherwise stated.

The columns were observed with a low-power microscope and freezing could easily be detected by movement and change of shape of the meniscus and by the opacity of the newly formed ice. The freezing points were measured with a mercury thermometer and are thought to be accurate to ± 0·2°c.

As a check that the capillaries had not leaked and admitted acetone, the melting points of all frozen specimens were found. Only a small percentage melted below 0°c, and the freezing points of these samples were rejected as unreliable.

3.5. *Experimental Results*

In table 1 the freezing points are tabulated according to the nature of the hydrophobic surface, and divided into 'long' and 'short' drops.

Table 1. The Freezing Point of Distilled Water in Hydrophobically Treated Glass Tubes

(1)	(2)	(3)	(4)	(5)	(6)		(7)
			Long drops				
Chlorosilane	12	0·25	68	−23·3	−28·5	−17·5	3·2
D 3032	15	0·25	44	−20·4	−22·3	−17·4	1·3
Drisil 29	12	0·24	62	−22·4	−23·9	−21·0	0·8
			Short drops				
Chlorosilane	12	0·25	2·6	−28·3	−34·7	−21·6	4·3
D 3032	15	0·25	1·3	−26·1	−31·9	−18·6	3·8
Drisil 29	12	0·24	1·4	−27·7	−33·2	−24·0	2·6

(1) Nature of hydrophobic coating; (2) number of tubes; (3) average diameter (mm); (4) average length of water (mm); (5) average freezing point (°c); (6) lowest/highest freezing points; (7) standard deviation.

Comparison with the results of Hosler and Hosler for ¼-mm tubes (see figure 1) shows that the present freezing points for long drops are much higher, presumably owing to the presence of freezing nuclei. The present results also disagree in showing a marked decrease in freezing point with droplet volume.

To find the reason for this discrepancy and also to locate the source of the impurities apparently present in our water samples, similar experiments were carried out using glass tubes that had not been made hydrophobic.

§ 4. EXPERIMENTS ON FREEZING OF WATER IN GLASS CAPILLARIES

4.1. *Preparation of Water Samples*

Capillaries of approximately 0·25 mm bore were drawn from Pyrex. Some were immediately filled with the water to be experimented on, others were first washed with strong chromic acid and distilled water.

The first experiments were performed with water distilled as described above. The effect of adding a small quantity of potassium permanganate to the water in the boiler was then explored. Finally some samples of water were made by low-pressure distillation from ice to a container cooled with liquid oxygen.

The technique of filling and sealing was the same as before. However it was not possible to observe the behaviour of a long and a short thread of water in the same capillary. In all cases where this was attempted it was found that the drops froze simultaneously (as closely as could be judged by eye). This was ascribed to nucleation of one drop by the other through the freezing of a thin film of water covering the walls between the two. Such cases were observed at distances of up to 5 cm between the drops.

It was therefore necessary to have only one drop in each tube, either a long (approximately 7 cm) or a short one (approximately 2 mm). To obtain an unambiguous comparison the long and short specimens were made alternately.

4.2. *Experimental Results*

Freezing points were measured as before and the results are shown in table 2.

The fact that these freezing points remained consistently high, as compared with the results of the previous section, and of Hosler and Hosler, no matter how carefully the water was distilled, indicated that the glass tubes themselves were contaminated with freezing nuclei.

Table 2. The Freezing Point of Distilled Water in Glass Capillaries

Category	(1)	(2)	(3)	(4)	(5)		(6)
All samples	141		0·25	−21·7	−31·6	−13·9	
All long drops	85	6·9		−21·5	−31·6	−14·5	
All short drops	56	0·17		−21·9	−29·4	−13·9	
Long drops filled	32	6·7	0·25	−21·1	−26·4	−14·5	2·0
Short drops alternately	34	0·17	0·25	−21·8	−29·4	−13·9	3·6
Tubes treated with chromic acid	94			−21·8			
Tubes not treated with chromic acid	47			−21·4			
$KMnO_4$ in still	42			−22·7			
No $KMnO_4$ in still	99			−21·2			
Water distilled at low pressure	34			−21·4			
Water distilled at ordinary pressures	107			−21·8			

(1) Number; (2) average length of water column (cm); (3) average diameter (mm); (4) average freezing temperature (°c); (5) lowest/highest freezing temperatures; (6) standard deviation.

Qualitatively, these results confirm the Hoslers' conclusion that the freezing points of water drops in glass tubes are virtually independent of drop length. This is in marked contrast to the behaviour of drops in hydrophobic tubes as reported above.

The clue to a possible explanation of this difference in behaviour lies in the observation that the walls of cleaned glass tubes containing water are apparently covered (or partially covered) by a thin film of water, whereas this will certainly not be the case when the walls have a hydrophobic coating. Glass capillaries drawn in the contaminated air of the laboratory, and cleaned as described above, appear to retain upon their walls freezing nuclei which may operate at temperatures as high as −14°c. Such particles, lodged upon the wetted wall will cause the freezing of the water film, and of any water droplet in contact with it.

The correctness of this theory has been verified by experiments described in the next section.

4.3. *Effect of altering Tube Length on Freezing Point of Water Droplet contained therein*

Sixteen Pyrex capillaries of approximately ¼ mm diameter were cleaned by soaking in chromic acid and then washing in distilled water. A single small drop of distilled water was then drawn into each tube and the ends sealed. The usual precautions, as described above, were taken to prevent room air entering the first nine tubes (Nos. 501 to 509). In the remaining tubes (510 to 516) room air was deliberately sucked in.

After the freezing points had been determined, the glass tubes were shortened by drawing them off in a flame about 2 cm away from the drops on either side. This was far enough away not to warm or disturb the drops. In three cases (Nos. 512, 514, 516) the effective (supercooled) lengths of the tubes were reduced instead by allowing about 9 cm to project above the level of the acetone in the cooling bath.

The freezing points were then measured again. In most cases they were measured twice to give some idea of the variations to be expected at constant tube length. An examination of the results set out in table 3 shows that reducing the effective length of the capillary either left the freezing point virtually unaltered (6 cases), or caused it to fall (10 cases). The decrease was about 2°c for those tubes from which room air had been excluded, and more in the other cases. No systematic decrease could be found in the freezing points repeated at constant tube length.

Table 3. Effect of altering Tube Length on Freezing Point of Water contained therein

(1)	(2)	(3)	(4)	(5)	(6)
501	0·25	12·6	−21·2	3·7	−22·2, −22·8
502	0·15	12·7	−21·8	3·9	−23·6, −22·9
503	0·35	15·7	−20·5	4·4	−20·3, −21·0
504	0·2	13·2	−21·1	3·9	−22·9, −22·6
505	0·9	14·2	−20·4	5·4	−23·6, −22·2
506	0·6	15·5	−21·3	4·6	−23·0, −23·5
507	0·2	11·3	−22·5	4·5	−22·6, −22·7
508	1·0	15·6	−21·9	5·5	−22·4, −22·2
509	0·6	11·6	−22·7	4·2	−23·1, −22·7
510	0·7	13·1	−18·7, −15·1	5·3	−19·5, −19·7
511	0·1	13·9	−20·7, −20·5	4·2	−24·2
512	0·3	17·9	−17·1, −16·0	9·9*	−21·4, −21·5
513	0·1	12·0	−22·2, −11·8	3·8	−22·1, −22·7
514	0·25	15·3	−20·3, −19·9	7·5*	−20·8, −21·1
515	0·35	14·7	−18·7, −18·8	3·3	−19·9, −20·2
516	0·4	17·5	−23·3, −23·1	5·8*	−23·0, −23·6

* Remainder of tube projected above cooling liquid.
Average bore of tubes = 0·24 mm

(1) Tube number; (2) drop length (cm); (3) initial length of tube (cm); (4) freezing point (°c); (5) final length of tube (cm); (6) new freezing point (°c).

It thus appeared that shortening the capillary often removed nuclei which had formerly initiated freezing of the enclosed drop. This confirms the hypothesis that foreign particles lodged upon the wall of a 'cleaned' glass tube can nucleate a thin film of water covering the walls and so cause the freezing of bulk water at a considerable distance. If the freezing nuclei upon the walls are effective at higher temperatures than those in the water drops themselves, then they will usually decide the freezing point of any drop in the tube, irrespective of length. It is to be expected that glass tubes prepared in the same way in the same room will have roughly similar nucleus contents; hence our observation that the average freezing points of long and short drops are virtually the same. The freezing nuclei operative in the above experiments cannot apparently be removed from a freshly drawn glass tube by washing in chromic acid or water.

The results of Hosler and Hosler can thus be accounted for by the presence of freezing nuclei upon the walls of the glass capillaries. Such nuclei would be fewer the narrower the bore of the capillary and thus the smaller its surface area;

hence the decrease in freezing point with tube diameter. There is no need to invoke Weyl's hypothesis which thus receives no support from these experimental results.

We have already mentioned that the results of the previous section using hydrophobic tubes appeared to be affected by the presence of freezing nuclei in the water itself. The chance of any drop containing a freezing nucleus effective at a particular temperature would be in direct proportion to its volume, and thus large drops would on the average freeze at higher temperatures.

It thus became apparent that neither the distillation technique nor the method of cleaning the glass tubes was capable of producing water drops which froze without the agency of impurities. A new technique had therefore to be adopted.

§ 5. THE FREEZING OF WATER CONDENSED DIRECTLY INTO FRESHLY DRAWN GLASS CAPILLARIES

5.1. *Preparation of Samples*

A Pyrex flask was fitted with a side tube T which was drawn out into a capillary of about ¼ mm bore (see figure 2). Water was boiled in the flask and the aperture of the top tap was adjusted so that a steady stream of steam passed down the capillary and condensed a few cm from the open end.

After boiling had continued for a measured time the capillary was sealed at P_1 above the condensation line, and then drawn off and sealed at P_2 about 10 cm higher. A certain amount of condensate would thus be enclosed in the capillary depending on the time between sealing at P_1 and P_2. A new capillary was then drawn and the process of filling repeated. In this way a number of water samples were made after successively longer periods of distillation.

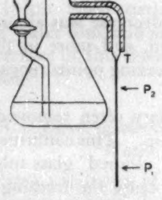


Figure 2. Still used for preparation of samples of water in glass capillaries.

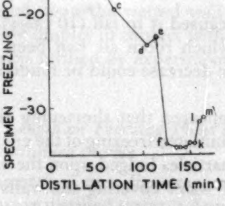


Figure 3. Freezing points of specimens of water prepared by condensation of steam after successively longer periods of distillation. (Typical example.)

5.2. *Freezing Points of Water Samples*

The freezing points of a batch of water samples prepared in this way showed certain reproducible phenomena, which will be described with reference to Batch l, figure 3.

The first capillary was drawn immediately after boiling commenced and was filled and sealed after a 15 minute steaming. This sample, a, had a freezing point of −13·5°c. Continued distillation produced specimens of successively lower freezing points until sample f, which froze at −33·9°c. Thereafter the freezing point remained constant within ±2°c for seven successive fillings at about 5 minute intervals.

All batches of water prepared in this way gave similarly decreasing freezing points, followed by a series at a more or less constant value $-32{\cdot}0 \pm 2{\cdot}5$°c. This suggests that the first specimens were contaminated by impurities which were eventually removed by the distillation process.

If impurities were driven out by prolonged boiling and the flask was left sealed and allowed to cool, then on restarting the distillation process, low freezing-point samples could be produced immediately. On the other hand, opening the flask and admitting room air invariably caused the freezing points to rise. Using side tube S, it was possible to admit distilled water to the flask without letting in any room air. The freezing points were then undisturbed.

From these facts it was concluded that the high freezing-point samples were contaminated by tiny solid particles originating in the room air; the discrete nature of these particles was shown by the fact that with prolonged boiling the freezing points changed, not smoothly, but in erratic steps.

To get 'constant' low freezing points it was thus essential to drive out all room air. This led to an improved technique in which the presence of room air was avoided by filling the flask completely with water which was then boiled away to give the steam for the water specimens.

A further precaution was found to be essential, namely that the capillaries should not be drawn until a stream of clean steam was passing through T. In almost every case where capillaries were made in room air, high freezing-point samples were produced; this despite soaking in chromic acid or steaming for as long as three hours before filling. It appears that some of the aerosol particles which become attached to the walls of a glass tube are immune to the conventional cleaning methods using chromic acid and steam. Their effectiveness as freezing nuclei apparently also survives heating to the softening point of Pyrex glass.

The drawing out of the wide tube T into capillaries served to spread out the freezing nuclei adhering to its walls, so that it was possible to make many successive samples without including any of them. But that their action was not completely destroyed was shown by the fact that samples of high freezing-point occasionally recurred, even when both the above precautions were observed.

Observations were made with as wide a range of tube diameters as was practicable. As the volume of water was increased above 10^{-3} cm^3 the proportion of obviously contaminated samples became steadily higher.

5.3. *Classification of Water Samples*

By examination of the curves of freezing point against distillation time, it was possible to divide the water samples into three classes:

(*a*) Those which contained freezing nuclei originating in the contaminated room air. They comprised the samples on the initial sloping portion of the graphs, isolated samples of much higher freezing points than their neighbours, and samples whose high freezing points were obviously a consequence of the admission of room air to the still. These will be called 'contaminated' samples (figure 3, samples a–e).

(*b*) Those specimens which occurred on the flat portion of the curve and which contained none of the freezing nuclei which could be removed by the precautions outlined above. It is not possible to be more specific at present about the agency responsible for their freezing, and they will be designated 'plateau' samples (figure 3, samples f–k).

(*c*) Those water samples which could not be assigned with certainty to either of the above categories and which will therefore be called 'unclassified' samples (figure 3, samples l, m).

Of 168 samples tested, 63 could be classed as 'contaminated', 89 as 'plateau' samples and the remaining 16 could not be classified. The frequency distribution of the freezing points in each category is shown in figure 4 (*a*).

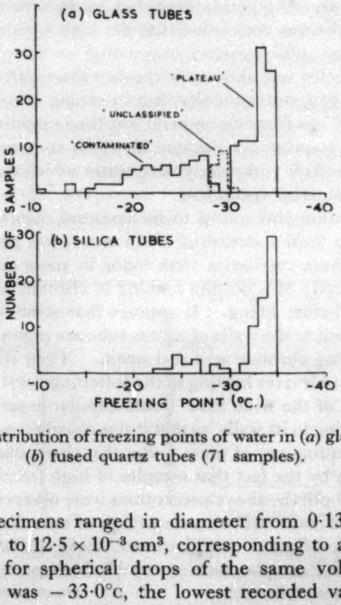


Figure 4. Frequency distribution of freezing points of water in (*a*) glass tubes (168 samples), (*b*) fused quartz tubes (71 samples).

The 'plateau' specimens ranged in diameter from 0·13 to 0·83 mm and in volume from 5×10^{-5} to $12{\cdot}5 \times 10^{-3}\,\mathrm{cm}^3$, corresponding to a range from 0·22 to 1·3 mm in diameter for spherical drops of the same volume. The average freezing temperature was $-33{\cdot}0$°c, the lowest recorded value being $-34{\cdot}5$°c. An examination of these data failed to show any clear relationship between the freezing point of a 'plateau' specimen and either its volume, diameter or surface area.

To enable present results to be compared with those of previous workers, the 'equivalent diameter' (i.e. the diameter of a sphere of equal volume) has been worked out for each 'plateau' and 'unclassified' sample and plotted against the corresponding freezing point. The areas in which these points lie are marked in figure 1.

5.4. *Effect of Cooling Rate on Freezing Point of Water Drops*

In an attempt to find possible differences in behaviour in water samples of the three categories, and also to check Bigg's observation (1953 a) that a reduction in cooling rate by a factor of 10, raises the average freezing point by approximately 2°C, a number of specimens were repeatedly frozen by cooling alternatively at rates of 0·5 deg min^{-1} and 0·05 deg min^{-1}. The freezing points are shown in table 4.

Table 4. Effect of Cooling Rate on Freezing Points of Water Drops

Freezing points are tabulated in °C for successive freezings at fast rate of cooling (0·5 deg min^{-1}) and at slow rate (0·05 deg min^{-1})

Spec. no.	Fast	Slow	Fast	Slow	Fast	Slow	Av. fast	Av. slow	Diff.
				' Plateau ' specimens					
329	−33·3	−33·5	−33·8	−33·6	−33·9	−33·6	−33·7	−33·6	0·1
355	−32·6	−32·9	−34·1	−33·7	−32·8	−33·9	−33·2	−33·5	−0·3
361	−31·5	−33·3	−32·8	−32·4	−32·0	−32·6	−32·1	−32·8	−0·7
370	−34·5	−32·1	−33·8	−33·9	−29·3	−28·2	−32·5	−31·4	1·1
372	−33·8	−32·1	−33·8	−33·7	−29·0	−28·1	−32·2	−31·3	0·9
				' Unclassified ' specimens					
303	−29·5	−29·2	−29·1	−28·6	−29·5	−28·6	−29·4	−28·8	0·6
323	−29·4	−27·0	−27·3	−26·0	−27·8	−28·4	−28·2	−27·1	1·1
353	−29·5	−28·9	−30·3	−31·9	−33·6	−28·9	−31·1	−29·9	1·2
				' Contaminated ' specimens					
334	−27·6	−27·0	−27·5	−26·9	−27·3	−26·0	−27·5	−26·6	0·9
403	−20·0	−20·1	−20·3	−20·2	−20·5	−20·4	−20·3	−20·2	0·1
409	−27·5	−29·5	−28·4	−27·0			−28·0	−28·3	−0·3
426	−28·8	−25·4	−26·0	−29·3	−31·3	−31·1	−28·7	−28·6	0·1

It will be seen that the differences in freezing point for the two rates of cooling were small and not very consistent. There were no marked differences in behaviour between the three categories of water sample. On the average there was a rise of 0·4°C in freezing point when the cooling rate was reduced by a factor of 10, though judging by the scatter of observations it is doubtful whether this difference is significant.

These conclusions are similar to those of Brewer and Palmer (1951) using water drops resting on a metal surface.

§ 6. THE FREEZING OF WATER IN SILICA TUBES

To investigate whether the freezing points of the 'plateau' samples was affected by the glass surface, similar experiments were carried out using fused silica in place of glass.

6.1. *Preparation of Samples*

A supply of 'clean' steam was produced by the improved technique described in the previous section, commencing with the still completely filled with water to avoid the presence of room air. A steady flow of steam was passed down a tube of transparent vitreous silica, 3 mm bore and 1 mm wall. From this, capillaries were drawn which were then partially filled with freshly condensed steam and sealed. The capillaries ranged in diameter from 0·1 to 0·27 mm. The difficulty of working silica prevented a larger range from being used.

6.2. *Freezing Points of these Samples*

The temperature distribution of the freezing points of 71 samples of water in silica is shown in figure 4 (*b*). It will be seen that they fall into two distinct groups similar to those into which the water-in-glass specimens were classified. Thus, by analogy, those freezing below -31°c may be called 'plateau' samples. It will be seen that the proportion of contaminated samples was greatly reduced by the new technique.

The lowest freezing point measured was $-34{\cdot}6^\circ$c. On plotting freezing points against 'equivalent diameter', it was found that all points for 'plateau' samples fell within the rectangle shown in figure 1. The range in volumes was from 0·05 to $3{\cdot}2 \times 10^{-3}$ cm^3, corresponding to a range in 'equivalent diameter' from 0·22 to 0·84 mm. There appeared to be no definite relationship between volume or surface area and freezing temperature.

The most striking feature of the results for 'plateau' samples of water in silica is their similarity to those in glass. With decreasing temperature the number of freezings rises steadily, until at a certain temperature, approximately the same for either glass or silica, all samples have frozen, further supercooling apparently being impossible.

§ 7. DISCUSSION OF RESULTS IN RELATION TO PREVIOUS WORK

From figure 1 it will be seen that the present freezing points (for both 'plateau' and 'unclassified' samples) are consistently lower than any previously reported mean freezing points of similar volumes of water. In default of any known means whereby the presence of solid supporting surfaces could hinder rather than assist nucleation in the present experiments, one must conclude that the results of previous workers were affected by the presence of freezing nuclei in the water used. This conclusion has already been reached in the case of Hosler and Hosler's results.

The present work has shown how difficult it is to free a glass apparatus of freezing nuclei in the form of solid particles that have entered in suspension in room air. It is possible that even the elaborate precautions of Wylie (1953) did not succeed in this purpose, and that some of his comparatively large samples were nucleated by such impurities and not by crystalline modifications of the glass as he suggests.

The earlier experimental results of Bigg (1953 a) show that the freezing of a water drop was a probability event dependent upon its volume and the rate of cooling. He concludes that freezing was either due to tiny motes (freezing nuclei) distributed throughout the water used, or else due to spontaneous nucleation. In his later paper (Bigg 1953 b) he adopts the latter explanation, since he finds that distilled water samples from different sources give similar freezing points. This may, however, show that such samples contain similar distributions of freezing nuclei which can occur even in carefully distilled water, as we have seen above.

The results of the present investigation indicate that Biggs drops larger than 50μ diameter did in fact contain freezing nuclei, since water samples of much lower freezing points for corresponding volume can be consistently prepared. The difference of approximately 10°c between the present lowest freezing points of 1 mm drops and Biggs' mean value for the same volume is much too large to be explained by the statistical spread of his freezing points about the mean, which amounts to only about $\pm 2°$c.

It is significant that Biggs' later observations on drops less than 50μ diameter lie below his previous curve. These may represent cases of spontaneous nucleation, whereas the weight of evidence indicates that his results on larger drops do not.

Neither of Biggs' suggested lines shown in figure 1 can therefore be accepted as the relationship between drop diameter and mean freezing point for spontaneous nucleation; nor should they be applied as such to the freezing of cloud droplets. An indication of the nature of the true relationship may be obtained from the lowest recorded freezing points of water drops of various sizes, as described in the next section.

In the present experiments the effect of cooling rate on freezing point was found to be much smaller than that discovered by Bigg. The discrepancy can only be resolved by further experimental work.

§ 8. Lowest Recorded Freezing Points

The lowest reliable freezing points for 'pure' water droplets of various sizes are of great interest, as they are more likely to represent cases of spontaneous freezing than any other available measurements. A search of the literature has yielded the following values which are plotted in figure 5.

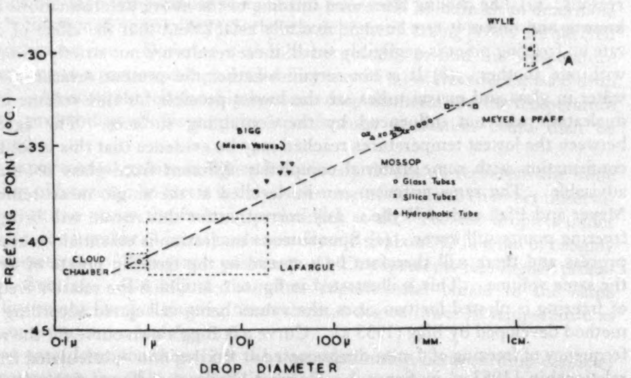


Figure 5. Relationship between diameter and the lowest recorded freezing points of water drops. Present measurements, and those of Meyer and Pfaff and Wylie are plotted according to 'equivalent diameters'. Dotted lines indicate estimated limits of accuracy.

(i) Observations on the '−40°c threshold' made by the present author with a small cloud chamber (described by Fournier d'Albe 1949). The first crystals were observed to form in a cloud of water drops condensed upon nuclei of various chemical compositions (e.g. $NaCl$, $MgCl_2$, $CaCl_2$), at approximately −40°c. With a concentration of 1000 drops per cm^3 (i.e. 100 in the field of view) almost all drops were observed to freeze if the cloud temperature fell below $-41{\cdot}2 \pm 0{\cdot}4$°c. Using the method of Wilson (1951) the duration of supersaturation in the chamber was estimated at 0·6 second, and the droplet diameter at 0·8 μ following Mason (1951).

(ii) The measurements of Lafargue (1950) mentioned above.

(iii) Bigg's observations on drops of diameter 15–50 μ. These have been included for completeness, though they represent mean values and not the lowest freezing points, which are unfortunately not available.

(iv) The present observations, which show good agreement between the lowest temperatures reached with water, in glass and fused silica. One freezing point of water in a tube with hydrophobic coating, at −34·7°c, has also been included.

(v) Meyer and Pfaff's (1935) value at −33°c. An estimate of the volume of this sample can be made. A tube 0·6 cm in diameter was used, and the sample was 'small'. If one takes the smallest depth of water that could be conveniently observed to freeze as 0·1 to 1 mm, this gives an equivalent diameter of 1·8 to 3·8 mm.

(vi) Wylie's (1953) lowest recorded freezing point of -30 ± 1°c. The volume of this sample was $2{\cdot}0 \pm 0{\cdot}7$ cm^3 (private communication).

In figure 5 line A has been drawn to show the most likely relationship on present evidence between drop diameter and lowest measured freezing point of water drops. This particular line cannot automatically be assumed to give accurate information on the spontaneous nucleation process for the following reasons: (*a*) The cooling rates used in some of the above determinations are not known, and unless it can be unequivocally established that the effect of cooling rate on freezing point is negligibly small, these results are not strictly comparable with one another. (*b*) It is not certain whether the present measurements on water in glass and quartz tubes are the lowest possible for this volume i.e. that nucleation was not influenced by the containing surfaces. The agreement between the lowest temperatures reached may be evidence that this is so, though confirmation with some material completely different from glass or quartz is advisable. The same criticism can be levelled at the single measurements of Meyer and Pfaff and of Wylie. Any correction for this reason will bring these freezing points still lower. (*c*) Spontaneous nucleation is essentially a statistical process and there will therefore be a spread in the freezing points of drops of the same volume. This is illustrated in figure 6 in which the relative frequency of freezing is plotted for two cases, the values being calculated according to the method developed by Bigg (1953 a). Curve *a* is Bigg's own curve for the relative frequency of freezing of 1-mm drops cooled at 0·5 deg min^{-1}, calculated from the relationship '1953 a' in figure 1. Curve *b* shows the shape of the frequency curve for the freezing of 1 mm drops cooled at 0·5 deg min^{-1} that would be obtained if line A, figure 5, represented temperatures for 50% probability of freezing of water drops by spontaneous nucleation. This assumption is made only to illustrate the nature of the resultant frequency distribution. The curves have been adjusted so that the maxima occur at the same temperature since we are interested at the moment rather in the shape of the curves, than in absolute temperatures, Curve *b* represents a far greater rate of increase of nucleation rate with falling temperature and this is reflected in the smaller spread of freezing points and sharper 'cut off' on the low temperature side. The similarity to figure 4(*b*) should be noted. From curve *b* it becomes obvious that the lowest measured value of freezing points of drops of a particular volume will depend on the number of drops measured, and isolated observations on single drops are thus of limited value.

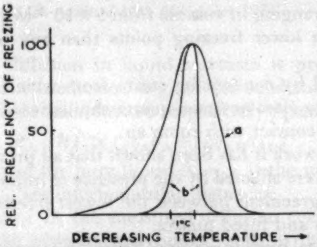


Figure 6. Relative frequency of freezing plotted against temperature for 1-mm water drops cooled at 0·5 deg min^{-1}, for two different relationships between droplet volume and the temperature for 50% probability of freezing : *a* that given by Bigg's line '1953 a' in figure 1, *b* that given by line A, figure 5.

A large number of measurements of the freezing point of high purity water drops of particular sizes in the diameter range 1 to 100μ are therefore required. A standard cooling rate, say 0·5 deg min^{-1} should be adopted. It will then be possible to plot the low-temperature wing of the frequency distribution curve. In practice the high temperature wing of this curve will be difficult to determine accurately because of uncertainty whether these higher freezing points are affected by impurities or not. However it should be possible to find the temperature corresponding to a particular high probability (say 95%) of freezing for each drop size used. An accurate relationship between drop diameter and the temperature for 95% probability of freezing by spontaneous nucleation could then be determined.

The experimental points in figure 5 obviously do not all represent the same probability of freezing; in fact it is only in the case of the cloud chamber observations that we can assess the freezing probability at all. For 100 droplets under observation the number of ice crystals rose from 1 to about 95 in a fall of roughly 1°c in temperature. The graph point at $-41{\cdot}2$°c represents this higher probability of freezing. To adjust the other experimental points to the same probability might involve moving them about a degree lower in temperature as we can see from curve *b* of figure 6.

These considerations show how the true relationship between drop diameter and temperature for 95% probability of freezing (say), at a cooling rate of 0·5 deg min^{-1}, may possibly differ from line A in figure 5. The possible differences will generally be in such a sense as to depress the freezing points of the droplets of diameter larger than 1μ still further.

From this we conclude that the effects of changing the droplet volume and cooling rate on the temperature of spontaneous freezing are much less than predicted by Bigg (1953 b). His suggestion that glaciation in cloud drops may occur spontaneously at temperatures up to $-20°$c, or even higher in large drops in cumulus cloud, is therefore invalid. The temperature at which spontaneous nucleation will become important in the freezing of cloud drops will be much lower than suggested by Bigg, though actual values can only be determined by further experimental work on the lines indicated above.

§ 8. CONCLUSION

Samples of water ranging in volume from 5×10^{-5} to 12×10^{-3} cm^3 have been prepared, which have lower freezing points than any previously reported for comparable volumes.

This was achieved by condensing steam from which all room air had been driven off, directly into glass or fused quartz capillaries. These capillaries were freshly drawn, out of contact with room air.

As a result of this work it has been shown that all previously reported results in this volume range were affected by the presence of impurities in the water.

There was close agreement between the lowest freezing points recorded for water samples in glass and fused quartz.

Further experimental work is required before an accurate relationship between droplet volume and spontaneous freezing temperature can be determined. However the available evidence is sufficient to show that spontaneous freezing will not play as important a part in the glaciation of clouds as suggested by Bigg.

ACKNOWLEDGMENTS

I wish to express my thanks to Professor Lord Cherwell for the use of the facilities of the Clarendon Laboratory. I am deeply indebted to Professor G. M. B. Dobson for continued advice and encouragement and to Dr. A. W. Brewer for helpful discussions. For the grant of a Research Scholarship I am indebted to the H. B. Webb Trustees, South Africa.

REFERENCES

BIGG, E. K., 1953 a, *Proc. Phys. Soc.* B, **66,** 688; 1953 b, *Quart. J. R. Met. Soc.*, **79,** 510.

BREWER, A. W., and PALMER, H. P., 1951, *Proc. Phys. Soc.* B, **64,** 765.

BRADLEY, R. S., 1951, *Quart. Rev. Chem. Soc.*, **5,** 315.

CWILONG, B. M., 1947, *Proc. Roy. Soc.* A, **190,** 137.

DORSCH, R. G., and HACKER, P., 1950, *N.A.C.A. Tech. Note* 2142, Washington.

FRENKEL, J., 1946, *Kinetic Theory of Liquids* (Oxford: University Press).

FOURNIER D'ALBE, E. M., 1949, *Quart. J. R. Met. Soc.*, **75,** 323.

HEVERLY, J. R., 1949, *Trans. Amer. Geophys. Un.*, **10,** 205.

HOSLER, C. L., and HOSLER, C. R., 1952, *Sci. Rep. No.* 1, Penn. State College, Contract No. AF 19(604)–140.

LAFARGUE, C., 1950, *Centenary Proc. R. Met. Soc.*, p. 61.

LA MER, V. K., 1952, *Industr. & Engng Chem.*, **44,** 1270.

MASON, B. J., 1951, *Proc. Phys. Soc.* B, **54,** 77; 1952, *Quart. J. R. Met. Soc.*, **78,** 22.

MEYER, J., and PFAFF, W., 1935, *Z. anorg. Chem.*, **224,** 305.

SCHAEFER, V. J., 1946, *Science*, **104,** 457.

VOLMER, M., 1939, *Kinetik der Phasenbildung* (Leipzig: Steinkopff).

WEYL, W. A., 1951, *J. Colloid Sci.*, **6,** 389.

WILSON, J. G., 1951, *Principles of Cloud-Chamber Technique* (Cambridge: University Press).

WYLIE, R. G., 1953, *Proc. Phys. Soc.* B, **66,** 241.

Mutual Viscosity and Diffusion in Liquid Mixtures

BY NORA E. HILL
Department of Physics, Bedford College, London

MS. received 16th November 1954

Abstract. A theory of diffusion in liquid mixtures is proposed which shows that the intrinsic diffusion coefficient, $\mathscr{D}_A$, of the substance A in a mixture of A and B in which the mole fractions are respectively f_A and f_B is

$$\mathscr{D}_A = \boldsymbol{k}T/6(f_A\eta_A\sigma_A + f_B\eta_{AB}\sigma_{AB})$$

where η_A is the viscosity of the pure liquid A, η_{AB} is the mutual viscosity of the mixture, and σ_A and σ_{AB} are intermolecular distances.

This leads to values of D, the measured diffusion coefficient, which are in good agreement with the experimental values, both in the limiting case when the concentration of one component is very low, and for 50% mixtures.

§ 1. INTRODUCTION

It has been shown previously (Hill 1954) that the dielectric relaxation time of a polar molecule in solution in a non-polar solvent is determined, not by the viscosity of the solvent, but by the mutual viscosity η_{AB} of the solute and solvent, which can be defined by

$$\eta\sigma = f_A{}^2\eta_A\sigma_A + f_B{}^2\eta_B\sigma_B + 2f_Af_B\eta_{AB}\sigma_{AB} \qquad \ldots\ldots(1)$$

where the f's are mole fractions, the σ's are intermolecular distances (calculated from $\sigma = (M/\boldsymbol{N}\rho)^{1/3}$, M = molecular weight, ρ = density, $\boldsymbol{N}$ = Avogadro's number) and the η's are viscosities. The suffixes A and B refer to the two kinds of molecule, and symbols without suffix refer to the mixture. The quantity $\eta_{AB}\sigma_{AB}$ can thus be calculated from experimental data. It is a constant for any pair of liquids at constant temperature.

The relation between relaxation time and mutual viscosity arises because the inner friction, which determines the relaxation time, is due almost entirely to interactions between molecules of different kinds. In an exactly similar way, the diffusion of molecules A into a solvent B is controlled almost entirely by the interaction of the two different kinds of molecule, when the concentration of A is low.

Assuming the diffusion process to be a consequence of the Brownian movement, the intrinsic diffusion constant, $\mathscr{D}$, as defined by Hartley and Crank (1949) is given by Einstein's equation

$$\mathscr{D} = \boldsymbol{k}T/\delta \qquad \ldots\ldots(2)$$

where $\boldsymbol{k}$ is Boltzmann's constant, T is the absolute temperature, and δ is the resistive force experienced by a molecule moving through the liquid with unit velocity.

§ 2. THE FRICTION CONSTANT

When the diffusing particle is large compared with the solvent molecules, it is permissible to use Stokes' equation

$$\delta = 6\pi\eta r \qquad \ldots\ldots(3)$$

where r is the radius of the diffusing particle, and η is the viscosity of the solution. But when the diffusing particle is of molecular dimensions this is no longer valid.

If, as in the case of dielectric relaxation, Andrade's model of a liquid is adopted (Andrade 1934), the resistive force arises from the loss of momentum which occurs when the molecule collides with its neighbours. Since there is no motion of the liquid as a whole relative to the coordinate system in terms of which the intrinsic diffusion coefficient is defined, the average drift velocity of the neighbours is zero, and the momentum interchange is the same as if they were at rest.

In a liquid in which the mole fractions of the two kinds of molecule are f_A and f_B, the loss of momentum from a molecule B to its neighbours per second is

$$f_B 4\nu_B c_B (m_B/2) v + f_A (2\nu_A + 2\nu_B) c_{AB} [m_A m_B/(m_A + m_B)] v$$

where ν_A and ν_B are the frequencies of vibration of the two kinds of molecule about their mean positions, c_A and c_{AB} are the probabilities of the two kinds of collision being effective in the transfer of momentum, and v is the velocity of the molecule under consideration.

Since

$$\eta_A = \frac{c_A}{3} \frac{\nu_A m_A}{\sigma_A} \qquad \ldots\ldots(4)$$

and

$$\eta_{AB} = \frac{c_{AB}}{3} \frac{(\nu_A + \nu_B)}{\sigma_{AB}} \frac{m_A m_B}{(m_A + m_B)} \qquad \ldots\ldots(5)$$

(Hill 1954) the loss of momentum per second is

$$6(f_B \eta_B \sigma_B + f_A \eta_{AB} \sigma_{AB}) v.$$

Therefore

$$\delta = 6(f_B \eta_B \sigma_B + f_A \eta_{AB} \sigma_{AB}). \qquad \ldots\ldots(6)$$

Similarly, the resistive force per unit velocity on a molecule of kind A is

$$\delta = 6(f_A \eta_A \sigma_A + f_B \eta_{AB} \sigma_{AB}). \qquad \ldots\ldots(7)$$

§ 3. THE DIFFUSION COEFFICIENT

Hartley and Crank (1949) have shown that, provided there is no appreciable change of the specific volumes of the two components on mixing, the measured diffusion coefficient is

$$D = \mathscr{D}_A V_B + \mathscr{D}_B V_A \qquad \ldots\ldots(8)$$

where V_A and V_B are the fractions of the unit volume occupied by A and B respectively.

From equation (7)

$$\mathscr{D}_A = \boldsymbol{k}T/6(f_A \eta_A \sigma_A + f_B \eta_{AB} \sigma_{AB}) \text{ etc.} \qquad \ldots\ldots(9)$$

therefore

$$D = \frac{\boldsymbol{k}T}{6} \left\{ \frac{V_B}{f_A \eta_A \sigma_A + f_B \eta_{AB} \sigma_{AB}} + \frac{V_A}{f_A \eta_{AB} \sigma_{AB} + f_B \eta_B \sigma_B} \right\}. \qquad \ldots\ldots(10)$$

In the limiting case when the concentration of one substance is very low, this becomes

$$D = \boldsymbol{k}T/6\eta_{AB}\sigma_{AB}. \qquad \ldots\ldots(11)$$

§ 4. COMPARISON WITH EXPERIMENTAL RESULTS

In table 1 the experimental and calculated values of the diffusion coefficient are presented, for the limiting case of very dilute solutions. The experimental values are taken from the sources shown, and the data for the calculation of the mutual viscosities from the *International Critical Tables*. In some cases extrapolation has been made to the temperature of the diffusion experiments on the assumption that the mutual viscosity varies with temperature in the same way as the viscosity of a 50 mole % mixture, or as one of the pure components. Since the correction is about 5–10%, this is not likely to introduce a serious error.

Table 1

Solvent	Solute	T°c	$10^{10}\eta_{AB}\sigma_{AB}$ poise cm	$10^5 D_{\mathrm{exp}}$ cm² sec⁻¹	$10^5 D_{\mathrm{calc}}$ cm² sec⁻¹	Ref.
$C_6H_5CO_2C_2H_5$	$CH_3CO_2C_2H_5$	16	6·1	0·76	1·09	1
CCl_4	$C_6H_5NO_2$	20	6·5	0·92	1·04	2
CCl_4	C_6H_6	20	4·1	1·40	1·64	2
C_6H_6	CCl_4	20	4·1	1·33	1·64	3
				1·89	1·64	2
C_6H_6	I_2	20	4·0	1·71	1·68	4
C_6H_6	C_6H_5Br	15	5·1	1·86	1·30	5
CH_3OH	$CHBr_3$	20	7·0	1·93	0·96	4
C_6H_6	$CHCl_3$	15	3·4	2·11	1·95	5
$CHCl_3$	I_2	20	3·2	2·12	2·10	4
C_6H_6	C_6H_5Cl	15	4·1	2·15	1·62	5
C_6H_6	$(C_2H_5)_2O$	15	1·9	2·21	3·0	5
C_6H_6	$C_2H_4Cl_2$	15	3·2	2·44	2·06	5
CS_2	I_2	20	3·2	3·10	2·10	4

1. Dummer 1919. 2. Gerlach 1931. 3. Munter 1931.
4. Stearn, Irish and Eyring 1940. 5. Thovert 1914.

The mixtures are placed in order of increasing value of the experimental diffusion coefficient. The calculated values are of the correct order of magnitude, and show a general increase as D_{exp} increases.

In table 2, the experimental results of Trevoy and Drickamer (1949 b) for hydrocarbon mixtures at 50% molar concentration are shown, together with the calculated values. The mutual viscosities are calculated from the viscosities and densities of the pure liquids taken from Timmerman's *Physico-Chemical Constants*, and the viscosities and densities of the 50% mixtures recorded by Trevoy and Drickamer (1949 a). In this case the agreement between the experimental and calculated values is very close indeed.

Table 2

A	B	T°c	$10^{10}\eta_{AB}\sigma_{AB}$ poise cm	$10^5 D_{\mathrm{exp}}$ cm² sec⁻¹	$10^5 D_{\mathrm{calc}}$ cm² sec⁻¹
C_6H_6	C_7H_{16}	25	2·04	2·47	2·79
C_6H_6	$C_{10}H_{22}$	25	3·80	1·72	1·79
C_6H_6	$C_{12}H_{26}$	25	5·12	1·40	1·44
C_6H_6	$C_{16}H_{34}$	25	7·70	0·96	1·06
C_7H_{16}	$C_{12}H_{26}$	25	4·58	1·58	1·56
C_7H_{16}	$C_{16}H_{34}$	25	5·26	1·00	1·34

§ 5. Conclusion

The theory is in closer agreement with experimental results than the theory of Stearn, Irish and Eyring (1940), which leads to values which are too high by factors varying from 2 to 5. It gives a more satisfactory account of Trevoy and Drickamer's results than Arnold's semi-empirical equation (Arnold 1905), which they found the most useful of existing formulae in correlating their results, since it does not involve any arbitrary constants.

Although the theory is here developed in terms of Andrade's model of a liquid, a similar argument will apply to any model which leads to a mutual viscosity term in the equation for the viscosity of a mixture.

References

Andrade, E. N. da C., 1934, *Phil. Mag.*, **17,** 497.
Arnold, J. H., 1905, *J. Amer. Chem. Soc.*, **17,** 549.
Dummer, E., 1949, *Z. anorg. Chem.*, **109,** 31.
Franke, G., 1932, *Ann. Phys., Lpz.*, **14,** 675.
Gerlach, B., 1931, *Ann. Phys., Lpz.*, **10,** 437.
Hartley, G. S., and Crank, J., 1949, *Trans. Faraday Soc.*, **45,** 801.
Hill, N. E., 1954, *Proc. Phys. Soc.* B, **67,** 149.
Munter, E., 1931, *Ann. Phys., Lpz.*, **11,** 558.
Stearn, A. E., Irish, E. M., and Eyring, H., 1940, *J. Phys. Chem.*, **44,** 981.
Thovert, J., 1914, *Ann. Phys., Lpz.*, **2,** 369.
Trevoy, D. J., and Drickamer, H. G., 1949 a, *J. Chem. Phys.*, **17,** 583; 1949 b, *Ibid.*, **17,** 1117.

Effects Produced by the Ionic Bombardment of Germanium

By W. D. CUSSINS
Department of Engineering, University of Cambridge

Communicated by C. W. Oatley ; MS. received 18*th August* 1954
and in final form 6*th January* 1955

Abstract. Specimens of pure single-crystal germanium and of polycrystalline germanium with varying amounts of impurity have been bombarded with positive ions of a number of elements having energies from 5 to 90 kev. The changes in some electrical properties of the specimen caused by this bombardment have been measured. In all cases it was found that maximum change was caused by a bombardment equivalent to about $6\mu c\ cm^{-2}$. No changes were observed with ion energies less than about 20 kev. The results are consistent with the assumption that bombardment produces a p-type layer whether the original specimen was p-type or n-type. The thickness of this layer has been measured for various energies of the incident ions.

---

§ 1. INTRODUCTION

THE investigation to be described was begun in 1950 and is somewhat similar to those described by Ohl (1952) and Kingsbury and Ohl (1952) on the ionic bombardment of silicon. Specimens of germanium containing differing amounts of impurity were bombarded with positive ions varying in atomic number from 1 (hydrogen) to 51 (antimony) and in energy from 5 kev to 90 kev. At the outset it was hoped that the ions might penetrate the surface layers of the germanium lattice and form interstitial impurity centres, but no evidence that this happens has been found. Various properties of the bombarded material have been measured, particularly its characteristics when used in a point-contact diode.

During the early experiments single-crystal germanium was not readily available and these investigations were made with polycrystalline material. In later experiments, single-crystal germanium was used.

§ 2. EXPERIMENTAL DETAILS

The apparatus used for bombarding the specimens is shown in figure 1. The ion source I in the case of gases was a Penning gauge with a small hole in one of the cathodes and for the metals was an arc discharge. The ion beam so produced was accelerated in three stages by electrodes E_1, E_2, E_3 and roughly focused at the specimen S, the beam current striking S being measured by the battery operated valve voltmeter. The magnet M was introduced to prevent loss of electrons from the specimen by secondary emission.

The power for E_1 and E_2 was supplied from a 50 c/s 3·5 kv transformer with voltage quadrupling rectifiers. A resistive tap across the output of the power pack supplied E_1. E_3 was taken from a small Van de Graaff type generator.

For measuring diode characteristics the specimen was mounted in a simple manipulator and the whisker was placed on a weighted balance arm to ensure constant contact pressure. The characteristic was then displayed on a calibrated oscilloscope from which tracings were taken. After a number of contacts had been made on a surface and all the characteristics traced on a single sheet of paper, the average characteristic was drawn by eye. No great difficulty was experienced in doing this.

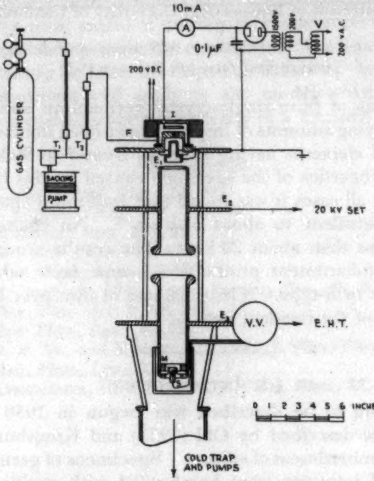


Figure 1. General arrangement of apparatus.

A standard method of preparing all specimens was adopted. They were first ground flat with a very fine grinding paste and then etched in 20 volume hydrogen peroxide for 15 minutes. The whiskers were of 0·1 mm tungsten pointed in an electrolytic etching bath of potassium hydroxide. Contact pressure was standardized at 0·25 g. No treatment for the specific purpose of forming the point contacts was used but, when measuring the diode characteristics, the applied alternating voltage was increased until the reverse 'turn-over' was observed, and this procedure produced some degree of forming.

§ 3. Changes in Diode Characteristics caused by Bombardment

In the first experiments polycrystalline germanium containing 150 parts per million (p.p.m.) of antimony was used. It was bombarded with 54 kev oxygen ions, with the results shown in figure 2 (*a*). It will be observed that the backward resistance falls sharply until a quantity of bombardment of 7·2 μc cm^{-2} is reached, after which further bombardment produced very little change. This general pattern of change was found to be the same for all types of germanium bombarded with all ions of any energy.

A detailed comparison of complete diode characteristics for a large number of specimens would have been very laborious, so the following simpler procedure was devised. Figure 2 (*b*) shows characteristics of various n-type specimens which have received identical bombardment. All these graphs together with those in figure 2 (*a*) have one feature in common, a sensibly linear portion of the reverse-current characteristic in the vicinity of the origin. Furthermore, to a first approximation, the forward current is not affected by bombardment. The current passed at a reverse bias of 4 v was therefore used as a convenient parameter for comparison. Plotting this against quantity of bombardment gives curves as in figure 3. In all experiments this type of curve resulted, the peak always occurring at about the same quantity of bombardment ($6\,\mu$c cm^{-2} which corresponds to only one ion for every 30 atoms in the surface layer of the germanium, assuming singly charged ions). The falling-off of current at large quantities of bombardment was attributed to specimen contamination. p-type specimens containing indium gave similar results. The changes in diode characteristic for a specimen bombarded with 78 kev oxygen ions are shown in figure 2 (*c*). Here there is a substantial decrease in both backward and forward resistances. As with n-type material the maximum effect is attained with a bombardment of $6\,\mu$c cm^{-2}.

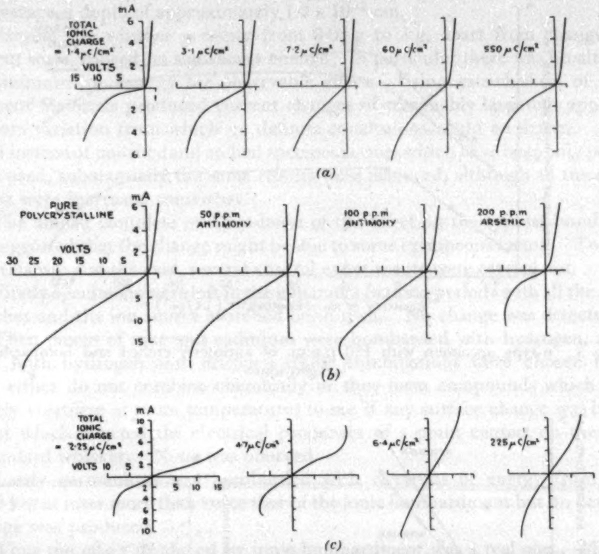


Figure 2. Diode characteristics of bombarded specimens. (*a*) n-type specimen with 150 p.p.m. of antimony, etched and bombarded with varying quantities of 54 kev oxygen ions. (*b*) n-type specimens with various impurities, etched and bombarded with 78 kev oxygen ions to $6\,\mu$c cm^{-2}. (*c*) p-type specimen with indium impurity, etched and bombarded with varying quantities of 78 kev oxygen ions.

Plotting the maximum change of reverse current at 4 v (obtained from curves like figure 3) against energy of bombarding ions gave curves similar to those in figure 4, which suggests a minimum ion energy of 20 kev to produce an effect. The maximum change is little affected by the quantity of impurity initially present in the germanium, although there is a tendency for the height to decrease with increasing impurity (figure 5). Nor is it greatly affected by the type of incident ion (ions of H, B, C, N, O, Al, A, Cu, Zn, Ge, Ag and Sb were tried) with two exceptions (B and Al) which could reasonably be ascribed to multiply charged ions. However, the heavier ions did tend to cause less effect.

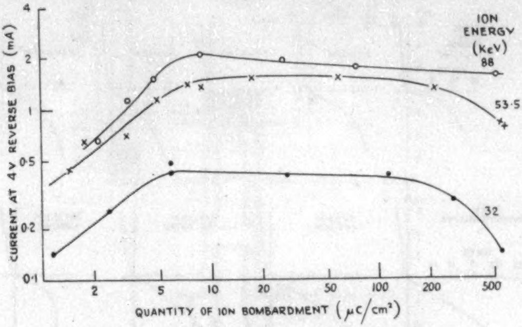


Figure 3. n-type specimen with 150 p.p.m. of antimony etched and bombarded with oxygen ions.

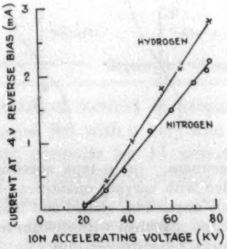


Figure 4. Maximum change in reverse current at 4 v for different ion energies.

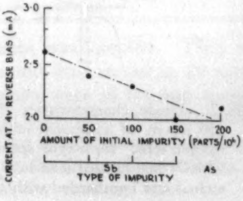


Figure 5. Maximum change in reverse current at 4 v for specimens with differing impurity, bombarded with 80 kev oxygen ions to 10 μc cm^{-2}.

The variation of effect with ion energy might have been due to an increased disturbance of the surface layers or the fact that higher energy ions affected a greater depth of germanium. To settle this point, bombarded specimens were etched in 20 volume hydrogen peroxide until their initial characteristics were restored, the time to do this being used as a measure of affected depth. This process, which gave very consistent results, revealed that an ion of given energy affected a definite depth, irrespective of the purity of the germanium and of the quantity of bombardment. It was also discovered that different ions accelerated by equal potentials affected the same depth with the same exceptions noted in the remarks on reverse current. These anomalous results support the hypothesis that the ions concerned were multiply charged, and they will therefore not be discussed further.

To make an estimate of the affected depth several specimens were etched for long periods and the amount removed was directly measured with a sensitive dial gauge. These results showed that etching for one minute removed a surface layer about 2×10^{-6} cm thick. This means that a 78 kev oxygen ion affects, on the average, a depth of approximately $1{\cdot}3 \times 10^{-5}$ cm.

Varying the whisker pressure from 0·05 g to 3 g, apart from changing the current scale, caused no significant change; in particular there was no alteration in minimum ion energy for observable effect. Using catswhiskers of several different materials produced current changes of reasonably large but apparently random variation from which no definite conclusions could be drawn.

If instead of polished and etched specimens, ones which have been only polished were used, substantially the same results were achieved, although all the current values were increased somewhat.

The almost complete independence of the effect on the type of bombarding ion suggested that the change might be due to some extraneous factor. To ensure that this was not the case, several control experiments were carried out.

Firstly specimens were left in the apparatus for long periods with all the voltage supplies and the ion source switched on in turn. No change was detected.

Then pieces of zinc and cadmium were bombarded with hydrogen, and tin with both hydrogen and nitrogen (these combinations were chosen because they either do not combine chemically or they form compounds which immediately volatilize at room temperature) to see if any surface change was brought about which affected the electrical properties of a point contact on them with a standard whisker. None was oberved.

Lastly germanium was bombarded with electrons of energy from 250 ev to 90 kev at rates more than twice that of the ionic bombardment but no detectable change was produced.

Thus the effect produced by ionic bombardment was a real one. Moreover it was found that, for a given type of ion, the change depended only on the total number of ions and on their energy. The rate of bombardment was unimportant and it made no difference whether a given amount of bombardment was given all at one time or in several separate stages. Tests over a period of one year showed that the changes produced were permanent and it was also found that the results could be reproduced with surprisingly good consistency.

At this stage of the work a statistical investigation of the spread of results obtained in a particular experiment was carried out. For a given surface, two hundred separate point contacts were made and the reverse current at 4 v measured for each. It was found that the standard deviations before and after bombardment were 10% and 7% respectively. Thus the mean value of forty results has a 95% probability of being within about 2·5% of the true value. This was considered good enough and each result quoted in this paper is the mean of 40 separate readings.

§ 4. FURTHER EXPERIMENTS WITH POLYCRYSTALLINE MATERIAL AND GENERAL REVIEW OF RESULTS

In addition to measurements of diode characteristics, thermoelectric measurements were also made on the surface of the specimens using the heated-whisker technique (Granville and Hogarth 1951). These showed without any ambiguity that, after bombardment exceeding $7\,\mu c\ cm^{-2}$ with ions having energies appreciably greater than, say, 30 kev, the surface layers were always p-type whether they had been n- or p-type before bombardment.

The effects of heating bombarded crystals in a vacuum for periods of about 30 minutes and then allowing them to cool slowly were also studied. It was found that the changes brought about by bombardment could be completely removed by heating in this way to temperatures above about 500°c, but that heating at lower temperatures produced no effect.

Taken together the foregoing results strongly suggest that the only effect of bombardment, over the range of energies that has been investigated, is the disruption of the surface layers of the germanium lattice to produce defects which can act as acceptor impurities (Lark-Horovitz 1951), and can be removed by annealing at 500°c. There is no evidence that any of the bombarding ions remained within the lattice as interstitial impurities since the effects observed were quite independent of the nature of the ion used.

It is interesting to compare these results with those obtained by Ohl (1952) for the bombardment of silicon. He obtained significant improvement in diode characteristics with ion energies as low as 1 kev but found that $600\,\mu c\ cm^{-2}$ were needed to produce maximum change.

§ 5. EXPERIMENTS WITH SINGLE CRYSTAL GERMANIUM

For the experiments to be described in this section specimens of n-type 'pure' single-crystal germanium of resistivity about 30 ohm cm were used. At the outset it was verified that the changes produced by bombardment on the thermoelectric and diode characteristics of this material were identical with those previously found with high-purity polycrystalline specimens, and that these changes could be removed by annealing at 500°c.

Specimens which had been bombarded with 78 kev ions of hydrogen, nitrogen or oxygen to a quantity of $6\,\mu c\ cm^{-2}$ were examined for transistor action, but no trace of such action was found. Voltage gains never exceeded 0·8 nor current gains 0·3 even when the curves were extrapolated to zero spacing between the contacts.

Measurements were also made of probe potentials on the surface near a point contact biased relatively to the base electrode. If E is the potential of the probe relative to the base at a distance b from the centre of the point contact when a current I is flowing, and if the surface of contact between the current-carrying point and the germanium is a hemisphere, then the quantity $2\pi bE/I$ should be independent of b and equal to the resistivity of the material (Brattain and Bardeen 1948). Before bombardment this was found to be the case even for large values of b, and also after bombardment if the contact were biased in the forward direction. With reverse bias however, abnormally high values of $2\pi bE/I$ were obtained, e.g. 70 ohm cm for material whose resistivity was 30 ohm cm.

§ 6. DISCUSSION

It has been shown that bombardment produces on the surface of the germanium a p-type layer whose depth is proportional to the energy of the incident ions. Furthermore, specimens of widely differing purity had substantially the same diode reverse-current characteristics after bombardment.

When a particular diode characteristic was displayed on the oscilloscope the alternating voltage across the crystal was gradually increased from zero until the reverse 'turn-over' of the characteristic was visible. During this process a certain amount of forming of the contact took place as was shown by changes in the shape of the low-voltage portion of the characteristic. If this is taken to indicate the occurrence of local melting of the germanium at the contact, it is reasonable to suppose that the point of the whisker will have penetrated the very thin p-layer on the surface of a bombarded crystal when the diode characteristic is being measured, and that the metal therefore makes separate contacts with the p-layer and the n-type material underneath. Such a model permits a simple explanation of the observed results, as follows. When the point is biased in the forward direction (with respect to the n-type material), practically all the current will flow through the n-type bulk of the specimen and the p-layer will have negligible effect. When however the point is biased in the reverse direction, little current can flow directly into n-type material because of the barrier set up at the interface. There is no such barrier between the metal and the p-layer, so current flows along this layer and thence to the base through the p–n junction of relatively large area. The p-layer thus causes a large increase in reverse current. Furthermore the p-layer is very thin and we shall see later that its conductivity is not very high, so that a large part of the applied reverse voltage is used in forcing current through the resistance of this layer. A detailed analysis of such a system suggests that the additional reverse current resulting from the p-layer should be roughly proportional to the applied voltage. This was found to be the case.

Let us assume then that the main features of the proposed model are correct. The thickness t of the p-layer is known, but not its resistivity ρ. However, the latter can be deduced from the probe-potential measurements with the point contact biased in the reverse direction, in the following manner. Let I be the current flowing radially outwards in the p-layer through a cylindrical surface of radius r and length t. Let V be the potential at this surface and let it decrease to $V-dV$ at radius $r+dr$. Then

$$dV = I\rho dr/2\pi rt \quad \text{or} \quad rdV/dr = I\rho/2\pi t. \qquad \ldots\ldots(1)$$

From the probe measurements V is known in terms of r so dV/dr can be found graphically and rdV/dr can be plotted as a function of r. By extrapolation we find the value of rdV/dr at the surface of the point contact where r is approximately equal to zero. On the assumption that the current flowing directly from the point into the n-region is negligible, the value of I at this surface is the total current flowing from the point. Substituting these values into equation (1), ρ can be calculated. In this way ρ was found to be 0·014 ohm cm, from which the density of ionized acceptors is calculated to be about 2×10^{17} cm^{-3}. This method is probably not very accurate but it should be right as to order of magnitude.

All the n-type polycrystalline specimens, as well as the pure single-crystal ones, gave very similar reverse-current curves after bombardment, and in all of them the thickness of the p-layer was the same for ions of the same energy. Since the reverse currents are negligibly small, in comparison, before bombardment, it may be concluded that in each case the p-layer contains about 2×10^{17} ionized acceptors when the maximum change has occurred. Before a p-layer can be formed however, electrons must be drained from the donors which were present originally. In the case of the single-crystal germanium the concentration of these donors was about 10^{14} cm^{-3}, whereas with the sample containing 200 parts per million of arsenic it must have been about 10^{19} cm^{-3}. It is extremely difficult to explain the fact that both of these samples give a p-layer with 2×10^{17} cm^{-3} ionized impurities after prolonged bombardment if we assume that the bombardment has produced only one type of defect. What is much more likely is that at least two types are produced and both can act as acceptors or traps for the electrons initially in the conduction band. Only those defects which give rise to acceptor levels close to the valence band will be ionized at room temperature and these probably constitute only a small fraction of the total. Thus the rate at which the initial electrons are drained away is much greater than that at which ionized acceptors are formed and most of the latter are formed after the disappearance of all electrons from the conduction band. Were this not the case the final number of ionized acceptors could hardly be the same in the two specimens under consideration. Somewhat similar conclusions have been reached by Brattain and Pearson (1950) as a result of their experiments on the bombardment of germanium with α-particles.

For the formation of defects by bombardment, two theories have in the past been considered (Lark-Horovitz 1951). On the one hand it is assumed that energy from the impinging particle is given up to certain atoms of the lattice which are separated from each other by other atoms that are not affected. In this way some of the lattice atoms will receive sufficient energy to cause them to move out of place and will thus form defects. It is difficult to reconcile the present results with this theory. In the first place only two types of defect can readily be envisaged—interstitial atoms and vacant sites—and of these the former would tend to act as donors (Lark-Horovitz 1951). But we have already seen the need to postulate at least two types of acceptor defects. Secondly, when the maximum change has been caused by bombardment, the number of acceptors in the p-layer is only 2×10^{17} cm^{-3}, or about one per million lattice atoms. Yet after this stage has been reached each bombarding ion must destroy as many acceptors as it creates. It is not easy to see how this can occur unless we postulate an extremely large effective cross section for the defects. Finally, there is the question of the thickness of the p-layer. Few measurements have been published of the penetration into solids of positive ions with energies in the range under consideration, but it seems unlikely that ions of oxygen and nitrogen could penetrate to the depths here recorded for the thicknesses of the p-layer. What is still more unlikely is that the depth of penetration for a given energy should be independent of the nature of the ion.

The alternative theory is that the energy of the ion is given up to a group of germanium atoms which are in proximity to each other so that, effectively, a small 'spot' of the crystal is melted. It is then assumed that solidification of this spot will occur so rapidly that a number of defects will be frozen in. In support of this theory it was formerly stated that thermal defects could be produced in bulk germanium by melting it and allowing it to solidify quickly, but more recent work (Slichter and Kolb 1952) appears to show that the defects in this case are caused by contamination with copper. However, the rate of cooling that can be achieved in macroscopic experiments of this kind is in no way comparable with that which would occur in the melted spot produced by the impact of an ion, when re-solidification might take place within a time of 10^{-12} second. There is therefore no fundamental objection to this theory and on other grounds it appears to fit the facts very well.

Using the data for bulk germanium given by Conwell (1952) the energy of an 80 kev ion should be sufficient to raise a volume of about 3×10^{-18} cm^3 of the material from room temperature to its melting point and to melt it. Such a volume would contain say 10^5 atoms and this is a large enough number to make the use of constants for bulk germanium not unreasonable. In all the experiments it was found that little further change was produced after the bombarding charge had reached 6μc cm^{-2}, and the volume of germanium melted by this quantity of 80 kev ions would be $1{\cdot}2 \times 10^{-4}$ cm^3, for an area of 1 cm^2. But the volume of the p-layer for the same area was found to be $1{\cdot}3 \times 10^{-5}$ cm^3. Apparently, therefore, the affected material has been melted about ten times. This factor is to be expected because the ions fall at random on to the surface and some regions will be melted several times before all have been melted once. Furthermore some of the incident energy will be conducted away by the lattice before it has contributed to the melting of a spot.

This theory provides a simple explanation of the fact that no further change occurs after a bombardment of 6μc cm^{-2}. Once the whole of the p-layer has been melted, additional melting will not increase the number of defects. The theory is also consistent with the fact that the depth of the p-layer is independent of the nature of the bombarding ion so long as its energy is the same. To explain the linear variation of depth with energy we must assume that the melted spot always has about the same average area of cross section parallel to the surface, and that the linear dimensions of this area are about 5×10^{-7} cm. This does not seem unreasonable, although somewhat different assumptions have been made in connection with theories of the sputtering of metals. In any case the accuracy of the present experiments was not very high and the range of energies investigated was only about fourfold. Hence the suggested linear variation of depth of penetration with energy may be an over-simplification of the truth. This however would not invalidate the other conclusions reached above.

The results reported here for germanium are quite different from those which Ohl obtained with silicon. The reason for this difference has already been discussed by Lark-Horovitz (1951) in another connection, and is to be found in the differing heights of the defect levels. It is assumed that in silicon these levels lie in the middle of the forbidden gap so that they are not ionized at room temperature.†

ACKNOWLEDGMENTS

The author is indebted to Mr. C. W. Oatley for suggesting the problem and for his continuous advice and encouragement throughout the investigation. He also wishes to thank the Director of the Research Laboratories of the General Electric Company for supplying the specimens of germanium.

The work was carried out in the Cambridge University Engineering Laboratories while the author was in receipt of a Department of Scientific and Industrial Research maintenance grant, and the cost of the apparatus was defrayed by a grant from the Admiralty.

---

† *Note added in proof.*—Since the present paper was submitted for publication, a paper by Lawrance, Gibson and Granville (1954) has appeared in which some results of the ionic bombardment of germanium are described. These authors, also, found that bombardment destroyed the rectifying properties of point contacts.

REFERENCES

BRATTAIN, W. H., and BARDEEN, J., 1948, *Phys. Rev.*, **74,** 231.
BRATTAIN, W. H., and PEARSON, G. L., 1950, *Phys. Rev.*, **80,** 846.
CONWELL, E. M., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 1327.
GRANVILLE, J. W., and HOGARTH, C. A., 1951, *Proc. Phys. Soc.* B, **64,** 488.
KINGSBURY, E. F., and OHL, R. S., 1952, *Bell Syst. Tech. J.*, **31,** 802.
LARK-HOROVITZ, K., 1951, *Semi-Conducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications).
LAWRANCE, R., GIBSON, A. F., and GRANVILLE, J. W., 1954, *Proc. Phys. Soc.* B, **67,** 625.
OHL, R. S., 1952, *Bell Syst. Tech. J.*, **31,** 104.
SLICHTER, W. P., and KOLB, E. D., 1952, *Phys. Rev.*, **87,** 527.

Radiation Induced Colour Centres in Fused Quartz

By M. LEVY and J. H. O. VARLEY
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

MS. received 7th July 1954 *and in amended form 20th December* 1954

Abstract. Samples of fused quartz ('Vitreosil') have been irradiated by pile, gamma or x-rays and three absorption bands are produced with maxima at 2·3 ev, 4·1 ev and approximately 5·6 ev. It is suggested that the 2·3 ev band is due to sodium borate impurities in the glass.

All three bands can be bleached either optically or thermally or, more rapidly, by a combination of both processes. An empirical law describing the thermal bleaching rate of the 2·3 ev band over a large range of n is $dn/dt = -Ae^{bn}$ where n is number of colour centres per cm^3 at time t; A, b are constants.

Bleaching of the 2·3 ev and 4·1 ev bands can also be brought about by large doses of pile irradiation. This is attributed to vacancies caused by atomic displacements combining with electron traps to produce new centres. These latter are assumed to give an absorption outside the range of observation in these experiments, viz. 6·2 ev to 1·2 ev.

§ 1. Introduction

Optical effects cannot be used to study radiation damage in metals and the variation of some other physical property, e.g. change in electrical resistance, must be used. This in general requires a large number of defects ($\sim 10^{18}$ cm^{-3}) to produce a noticeable change. The alkali halides and glasses however are much more sensitive to radiation damage as, according to Smakula's (1930) formula for F-centres, 10^{14} colour centres per cm^3 are sufficient to give a measurable absorption band.

Radiation and chemically induced absorption bands in the alkali halides and glasses have been extensively studied for many years. Pohl (1937) and his co-workers, Mott and Gurney (1940) and Seitz (1946) are notable in the field of the alkali halides, and their researches have indicated that the visible absorption bands are due to electrons trapped at lattice defects. Fröhlich (1947), Turner and Lewis (1947) and Simpson (1950) have also said that there are a large number of defects that can act as trapping centres in a glass. If fused quartz were a purely homopolar substance, then it should not be radiation coloured by the mechanisms usually associated with ionic materials such as the alkali halides and ordinary glasses. However, it has been observed by many workers that both quartz crystal and fused quartz may be coloured similarly to the alkali halides, therefore either the fused quartz has characteristics of the polar bond or impurities are responsible. In the discussion that follows it is suggested that the latter explanation is the most likely.

Several workers have observed the 2·3 ev, 4·1 ev and 5·6 ev absorption bands in fused quartz: Mayer and Guéron (1952) and Yokota (1952 a) by x-irradiation, Mitchell and Paige (1954) by neutron radiation and Dainton and Rowbottom (1954) by x- and γ-rays. Yokota also made fused quartz in an oxidizing atmosphere and only observed the 4·1 ev and 5·6 ev bands, i.e. no visible coloration; this case of no coloration being produced by radiation has also been observed by Gale and Bickford (1953) for electron irradiated specimens of Corning fused silica.

Mayer and Guéron (1952), Yokota (1952 a) and Dainton and Rowbottom (1954) have all observed that there is a relationship between the 2·3 ev and 5·6 ev absorption bands both in their increase in height with increasing radiation dosages and in their behaviour on optical or thermal bleaching. Mayer and Guéron and Dainton and Rowbottom have observed that the 4·1 ev band bleaches more quickly than the 2·3 ev and 5·6 ev bands on optical or thermal bleaching. Dainton and Rowbottom have also observed that the 4·1 ev band may grow at the expense of the 5·6 ev band on selective optical bleaching and vice versa. These authors have all reported the presence of a small absorption band at approximately 5·2 ev in unirradiated specimens which, according to Dainton and Rowbottom (1954), increases in size with increasing impurity of the specimen.

Concerning the radiation of quartz crystal, Mitchell and Paige (1954) have observed broad bands having maxima at 2·8 ev and 5·6 ev, whereas Yokota (1952 b) only observed a general absorption with no distinct absorption bands. Thomas *et al.* (1949) observed that synthetic quartz, either coloured purple or smoky brown, and again Brown and Thomas (1952) found that synthetic quartz, grown by the temperature gradient method to give a high purity crystal, did not colour at all on x-irradiation. The first indication that the colouring in quartz might be due to impurities was a result of the work of Brown and Thomas (1952) who found that the smoky colouring in quartz crystals sometimes followed the rhombohedral planes and this was thought to be due to the way the crystal was grown. Additions of copper, chromium, manganese and arsenic increased the rate of coloration through x-irradiation.

The object of this paper is to report upon the effects of radiation on fused quartz and in the light of the information obtained to offer a tentative explanation of the absorption bands observed together with suggestions for further work.

§ 2. EXPERIMENTAL

The glasses used throughout all experiments, except where otherwise stated, were 'Vitreosil' (fused quartz) optically flat discs 20 mm in diameter × 3 mm thick, manufactured by the Thermal Syndicate Co. Ltd.

A spectrographic analysis of the glass revealed that it contained about 0·4% by weight impurities, made up of 0·3% boron, 0·04% copper, 0·05% sodium, 0·005% iron.

Sodium borate and sodium silicate glasses were made from 'Analar' chemicals by melting in an oven and pouring the melt on to a steel plate, then flattening into discs 1 mm thick and about 15 mm in diameter. These were only irradiated in a gamma source and not in the pile, as this would have caused high activities and thus have delayed our observations. The pile-irradiated fused quartz needed a few days for its associated activities to die down but it has been observed that the absorption spectrum of the glass does not alter appreciably after storing it in the dark for a period of months.

Sources of radiation for colouring the glass specimens were either 40 kv 15 mA x-rays with the specimen 1½ inches from a copper target, a 100 curie cobalt 60 gamma source giving a flux of approximately $4{\cdot}5 \times 10^4$ r/hour, or Bepo pile radiation giving an integrated slow neutron flux of approximately 10^{12} n cm^{-2} sec^{-1}. All three sources of radiation gave rise to the same three absorption bands, the pile producing them the most rapidly due to the higher fluxes involved. It has been observed that 14 hours' x-irradiation as defined above, 10^6 r gamma irradiation or 2×10^{14} n cm^{-2} pile irradiation produce approximately the same intensities for the A, B and C bands. A high pressure Ediswan MBL/D mercury arc was used for bleaching experiments, with or without selective ultra-violet filtering.

Observations of absorption spectra were made with a Unicam SP.500 Spectrophotometer in the range 6·20 ev (200 mμ) to 1·25 ev (1000 mμ). Absorption spectra have been recorded in terms of absorption coefficient as a function of the photon energy and all readings have been corrected for reflection losses.

§ 3. Results

The coloration induced in all fused quartz specimens was the same patchy purple for x-rays, gamma rays or pile irradiation, the patchy effect increasing on further time of irradiation. Figure 1 shows the absorption spectrum of a fused quartz sample both before and after $2{\cdot}5 \times 10^6$ r gamma irradiation. In the unirradiated curve there is a small absorption band at 5·14 ev and indication of a general increasing absorption in the far ultra-violet region. The irradiated curve shows three absorption bands, each about two electron volts wide, with maxima situated at 2·3 ev, 4·1 ev and about 5·6 ev. These will in future be referred to as the A, B and C bands respectively and the associated centres the A, B and C centres.

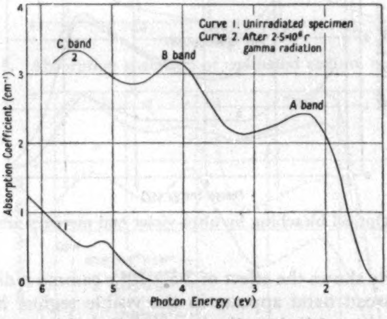


Figure 1. Absorption spectrum of radiated fused quartz.

Figure 2 shows the effect on the absorption spectrum of optical bleaching by light of photon energies smaller than 4·0 ev obtained by using a mercury arc with a hysil glass filter. The B band diminishes the most rapidly and the A and C bands appear to decrease at the same rate. Figure 3 shows how the same effect is obtained with ultra-violet light (3·5–5·0 ev) bleaching alone and also shows how unfiltered mercury arc bleaching causes the whole absorption spectrum to decrease until the initial unirradiated absorption spectrum is obtained.

Figure 4 shows the effect of 5×10^6 r gamma irradiation on a sodium tetraborate glass. The absorption edge or strong absorption band at about 5·2 ev appears to move to about 4·4 ev and an absorption band at 2·3 ev is observed; this is the same shape and at the same position as the A band in irradiated fused quartz.

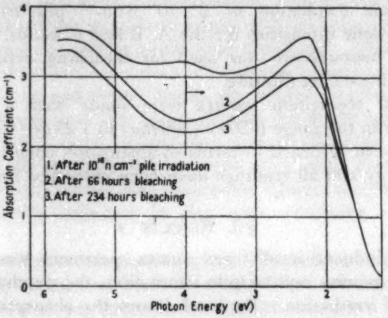


Figure 2. Optical bleaching by light of photon energies less than 4 ev

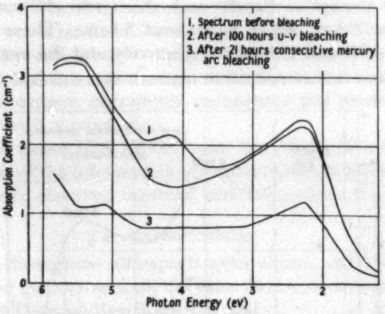


Figure 3. Optical bleaching by ultra-violet and mercury arc radiation.

Figure 5 similarly shows the effect of $2{\cdot}5 \times 10^6$ r gamma radiation on a sodium silicate glass. A broad band appears in the visible region, having no distinct maximum and merging with the tail of general absorption in the ultra-violet.

Figure 6 shows the effect of increasing pile irradiation on the absorption spectrum of a sample of fused quartz. The A and B bands appear to reach a maximum and then begin to decrease and thus the glass itself becomes less densely coloured. It is difficult to judge the behaviour of the C band as there is a marked increase in μ in this region. The apparent shift of the C band maximum to higher energies is thought to be due to a shift of a general absorption towards lower energies which after about $2 \times 10^{18}\,\mathrm{n\,cm^{-2}}$ almost reaches the visible range. This is thought to be due to the absorption of the SiO_2 itself, initially

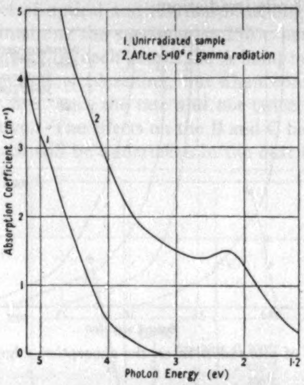


Figure 4. Absorption spectrum of irradiated sodium tetraborate.

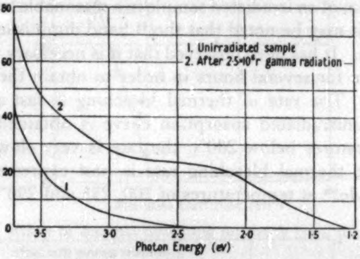


Figure 5. Absorption spectrum of irradiated sodium metasilicate.

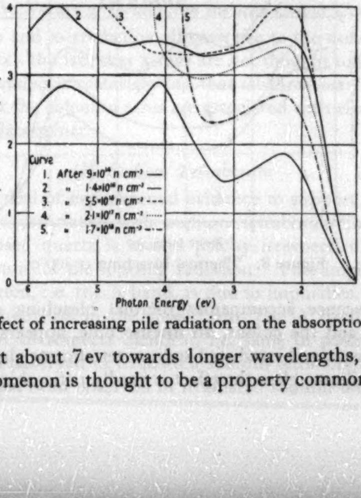


Figure 6. Effect of increasing pile radiation on the absorption spectrum.

outside the range at about 7 ev towards longer wavelengths, i.e. lower photon energies; the phenomenon is thought to be a property common to the homopolar bond. Figure 7, showing the effect of increasing pile irradiation on a homopolar acrylate, illustrates how a strong absorption band moves from 4·2 ev to 3·0 ev after nearly 10^{17} n cm^{-2} pile irradiation.

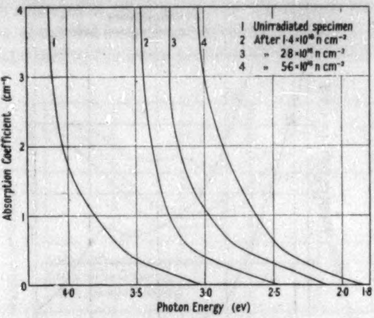


Figure 7. Effect of pile radiation on the absorption edge of a co-polymer.

Thermal bleaching at about 300°c brings about measurable changes in the absorption spectrum of an irradiated sample in a reasonable time, this is illustrated in figure 8 where it may be noted that the B band diminishes more rapidly than the A and C bands. It has been observed that it is necessary to maintain the glass at this temperature for several hours in order to obtain the initial unirradiated absorption curve. The rate of thermal bleaching is fast above 300°c, and at 500°c the initial unirradiated absorption curve is obtained in a few minutes, whereas at temperatures below 200°c the rate is very slow indeed. Figure 9 illustrates how the thermal bleaching rate is best expressed by the empirical relation $dn/dt = -Ae^{bn}$ at temperatures of 200, 255 and 290°c for three different specimens.

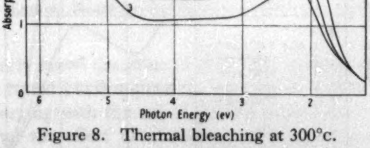


Figure 8. Thermal bleaching at 300°c.

Thermoluminescence accompanies thermal bleaching of irradiated fused quartz specimens, and the colour, an intense blue, appears the same for both irradiated fused quartz and irradiated sodium tetraborate glasses, whereas in the irradiated sodium silicate glass the effect is hardly noticeable. This effect of the thermoluminescence of irradiated fused quartz has also been observed by Yokota (1952 a) and Dainton and Rowbottom (1954).

The combined effect of optical and thermal bleaching on the A band using a mercury arc and maintaining the specimen at 250°c has been studied. Such optical or thermal bleaching by itself causes the A band to decrease by the same amount for the same period of bleaching, but when combined they cause the A band to diminish at five times the rate and not twice as would be expected if the effects were additive. The effects on the B and C bands have not yet been studied and such research will be undertaken in the near future.

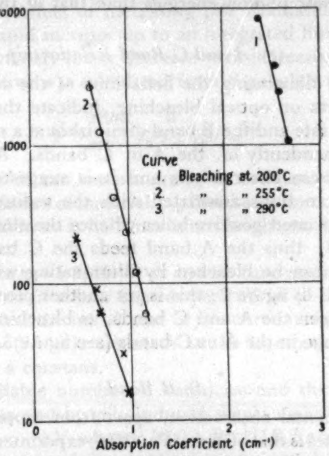


Figure 9. Effect of thermal bleaching on the A band maximum.

§ 4. Discussion

(*a*) *The Nature of Irradiation induced Colour Centres*

Neutron, gamma and x-irradiation all gave rise to the same three absorption bands in fused quartz; the fact that x-rays are not thought to be able to produce displacements in homopolar materials and that all three bands can be bleached optically implies that the colour centres are associated with electronic excitations and not atomic displacements.

(*b*) *Sodium Tetraborate*

There is a great deal of experimental evidence to support the interpretation that the A band is associated with sodium tetraborate impurity: (i) The coloration of the fused quartz is always patchy irrespective of the particular sample or of the nature of the ionizing radiation. This implies that the cause of the visible coloration, i.e. the A band, is due to impurities. (ii) Sodium and boron are known impurities in the glass. (iii) The A band in the irradiated glass samples was in all respects identical, i.e. same in position and shape, as the absorption band observed in irradiated sodium tetraborate. An absorption band at 2·3 ev has also been observed in irradiated sodium tetraborate by Sojka bond. Figure 7, showing the effect of increasing pile irradiation on a homopolar acrylate, illustrates how a strong absorption band moves from 4·2 ev to 3·0 ev after nearly 10^{17} n cm^{-2} pile irradiation.

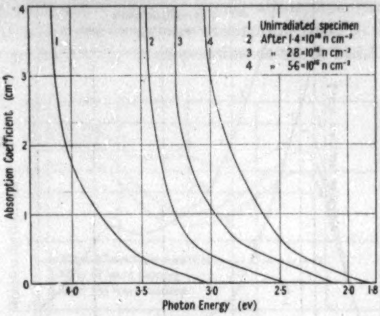


Figure 7. Effect of pile radiation on the absorption edge of a co-polymer.

Thermal bleaching at about 300°c brings about measurable changes in the absorption spectrum of an irradiated sample in a reasonable time, this is illustrated in figure 8 where it may be noted that the B band diminishes more rapidly than the A and C bands. It has been observed that it is necessary to maintain the glass at this temperature for several hours in order to obtain the initial unirradiated absorption curve. The rate of thermal bleaching is fast above 300°c, and at 500°c the initial unirradiated absorption curve is obtained in a few minutes, whereas at temperatures below 200°c the rate is very slow indeed. Figure 9 illustrates how the thermal bleaching rate is best expressed by the empirical relation $dn/dt = -Ae^{bn}$ at temperatures of 200, 255 and 290°c for three different specimens.

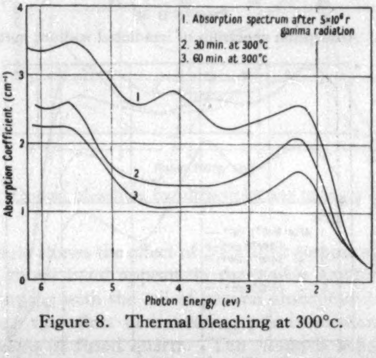


Figure 8. Thermal bleaching at 300°c.

Thermoluminescence accompanies thermal bleaching of irradiated fused quartz specimens, and the colour, an intense blue, appears the same for both irradiated fused quartz and irradiated sodium tetraborate glasses, whereas in the irradiated sodium silicate glass the effect is hardly noticeable. This effect of the thermoluminescence of irradiated fused quartz has also been observed by Yokota (1952 a) and Dainton and Rowbottom (1954).

The combined effect of optical and thermal bleaching on the A band using a mercury arc and maintaining the specimen at 250°c has been studied. Such optical or thermal bleaching by itself causes the A band to decrease by the same amount for the same period of bleaching, but when combined they cause the A band to diminish at five times the rate and not twice as would be expected if the effects were additive. The effects on the B and C bands have not yet been studied and such research will be undertaken in the near future.

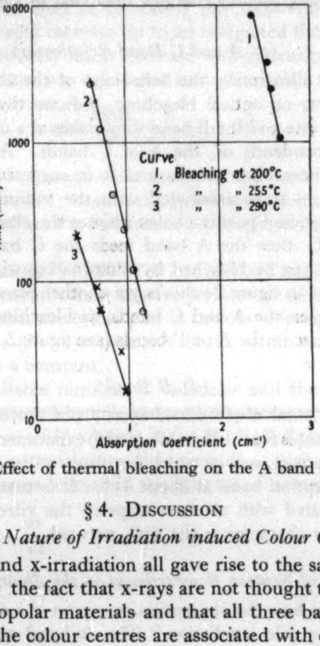


Figure 9. Effect of thermal bleaching on the A band maximum.

§ 4. Discussion

(*a*) *The Nature of Irradiation induced Colour Centres*

Neutron, gamma and x-irradiation all gave rise to the same three absorption bands in fused quartz; the fact that x-rays are not thought to be able to produce displacements in homopolar materials and that all three bands can be bleached optically implies that the colour centres are associated with electronic excitations and not atomic displacements.

(*b*) *Sodium Tetraborate*

There is a great deal of experimental evidence to support the interpretation that the A band is associated with sodium tetraborate impurity: (i) The coloration of the fused quartz is always patchy irrespective of the particular sample or of the nature of the ionizing radiation. This implies that the cause of the visible coloration, i.e. the A band, is due to impurities. (ii) Sodium and boron are known impurities in the glass. (iii) The A band in the irradiated glass samples was in all respects identical, i.e. same in position and shape, as the absorption band observed in irradiated sodium tetraborate. An absorption band at 2·3 ev has also been observed in irradiated sodium tetraborate by Sojka

(1940) and in borax by Urbanek (1951). (iv) The irradiated sodium tetraborate glass and fused quartz glass exhibit the same colour on thermoluminescence. (v) The strong absorption in the unirradiated sodium tetraborate glass at about 5·2 ev corresponds to the same position as the small absorption band in specimens of unirradiated fused quartz, see figure 1.

The evidence against associating the A band with sodium silicate impurities is that the irradiated sodium silicate glass exhibited neither an absorption band at 2·3 ev nor such an intense thermoluminescence as the sodium tetraborate. In addition the strong absorption of the unirradiated sodium silicate glass occurred at much lower photon energies than that of the unirradiated sodium tetraborate.

(*c*) *A and C Band Relationship*

Figures 2 and 3, illustrating the behaviour of the absorption spectrum of irradiated fused quartz on optical bleaching, indicate that the A and C bands diminish at the same rate and the B band diminishes at a much greater rate than, and apparently independently of, the A or C bands. As reported in §1, the literature supports these observations and it is suggested that the A centres arise from electrons in traps associated with the sodium tetraborate and the C centres are the associated positive holes whence the electrons in the A centres were initially ejected; thus the A band feeds the C band on bleaching. In addition the C band can be bleached by illuminating with light in the A and B bands, as illustrated in figure 2; this is yet another piece of evidence showing the relationship between the A and C bands, as bleaching in the B band alone causes very little change in the A or C bands (see figure 3, curve 2).

(*d*) *B Band*

The only experimental observation, apart from those mentioned in §1, of the 4·1 ev band in glasses is that of Yokota's (1954) experiments of the x-irradiation of various alkali silicate glasses in which he observed in about twelve different glasses a strong absorption band at about 4 ev. It is tentatively suggested that the B band is associated with electron traps in the vitreous quartz, probably with a degree of polar character in the SiO_2 network.

(*e*) *The Effect of Neutron Bombardment on the Absorption Spectrum*

Figure 6 illustrates how the absorption spectrum of a sample of fused quartz changes on increasing pile irradiation. The A band reaches a maximum at 4×10^{16} n cm^{-2} and the B band reaches a maximum at $2{\cdot}5 \times 10^{16}$ n cm^{-2}; thereafter both decrease with increasing time of irradiation. The effect on the C band is not known as it is masked by a movement towards longer wavelengths of a strong absorption which is possibly associated with the silica network.

It is suggested that the bleaching of the A and B bands is due to a combination of pile created vacancies with A and B centres to give new centres whose absorptions lie outside the observable range, viz. 6·2–1·2 ev. There are grounds here for further research as the suggestion for the destruction of the A and B centres is only tentative. It would also be of advantage to study the effects of increasing x- and gamma irradiation on the absorption spectrum to see whether there is any decrease in the intensities of the absorption bands on increasing doses of these irradiations.

The observation that the absorption edge in homopolar organic compounds moves towards longer wavelengths on increasing pile irradiation is well known and has been observed in certain plastics by Monk (1952) and Best (1953). In figure 7 an example of this is illustrated in the case of a co-polymer where the strong absorption band or edge moves one electron volt after $5{\cdot}6\times10^{16}$ n cm^{-2} irradiation. It is feasible that the same effect could occur in fused quartz which is generally assumed to be a homopolar substance.

(*f*) *Destruction of A Centres by Increasing Pile Irradiation*

Figure 10 shows the effect of increasing pile irradiation on the number of A centres; there is a rapid increase up to an integrated flux of $0{\cdot}5\times10^{17}$ n cm^{-2} followed by an approximately linear decrease with increasing irradiation dose.

The number of centres was calculated from Smakula's formula:

$$nf=\frac{18mc}{\pi e^2h}\frac{n'}{(n'^2+2)^2}\mu_{\mathrm{m}}W$$

where n is the number of centres per cm^3, f is the oscillator strength of each centre, n' the refractive index of material, μ_{m} the absorption coefficient at the peak, W the width of the band in ev at half maximum height, m, h, c, e have their usual meanings.

For the vitreosil glass $n'=1{\cdot}5$, $W=1$ ev and we have assumed $f=1$. Therefore $n\sim2\mu_{\mathrm{m}}$. ($n=10^{-17}\times$ no. of colour centres per cm^3 in all equations.)

It has been found that the initial part of the curve in figure 10 approximates to the equation $dn/dt=(N-n)a$ where N is the number of available negative ion vacancies, and a is a constant.

Now both the available number of vacancies and the A centres produced are assumed to be destroyed by pile-induced displacements. Therefore N is a function of t and $dN/dt=-bN$, whence $N(t)=B\mathrm{e}^{-bt}$ (B, b are constants). Further, dn', the number of A centres destroyed in time dt, is such that $dn'/dt=-cn$, so that the number of centres formed in the interval of time dt is

$$\Delta n=\frac{dn}{dt}dt+\frac{dn'}{dt}dt$$

$$=\{N(t)-n\}a\,dt-cn\,dt.$$

Hence $$\frac{\Delta n}{dt}=aN(t)-(a+c)n=aB\mathrm{e}^{-bt}-(a+c)n$$

and $$\frac{\Delta^2n}{dt^2}=-abB\mathrm{e}^{-bt}-(a+c)\frac{\Delta n}{dt}=-b\frac{\Delta n}{dt}-b(a+c)n-(a+c)\frac{\Delta n}{dt}.$$

This has a general solution $n=R\mathrm{e}^{rt}+S\mathrm{e}^{st}$, where R, r, S, s are all constants. Choice of suitable values of these parameters gives good agreement with the curve in figure 10.

In calculating the effect of displacement formation on the number of A centres it is assumed that all the sodium is present as sodium tetraborate, the remaining boron being distributed throughout the glass at random. The sodium tetraborate is assumed to be segregated, as is inferred from the patchy coloration resulting from irradiation.

The fission fragments from the reaction $^{10}_{5}\mathrm{B} + n = ^{7}_{3}\mathrm{Li} + ^{4}_{2}\alpha$ have been calculated to cause approximately tenfold the displacement damage in the sodium tetraborate networks as the fast neutrons themselves cause.

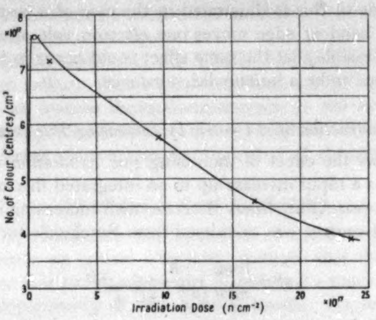


Figure 10. Bleaching of A centres by pile radiation.

Using the equations derived by Seitz (1949) for the proportion of energy of a fast particle, viz. $^{7}_{3}\mathrm{Li}$, in causing displacements, and assuming that fission fragments only cause displacements in the region where they are formed, i.e. in the sodium tetraborate networks, an interesting result can be derived.

If we assume that the $\mathrm{Na_2B_4O_7}$ networks consist of a monatomic lattice of atom $^{15\cdot5}_{7\cdot5}Y$ with a fast neutron cross section of 6 barns, we can show that as the number of Y atoms per cm^3 $\sim 10^{20}$, the number of primary collisions is of the order of 5×10^{13} per day, assuming a fast neutron flux of 10^{12} cm^{-2} sec^{-1}.

Assuming the displacement energy to be 25 ev, each primary collision results in approximately 63 displacements and so the fast neutrons cause approximately 4×10^{15} displacements per day in the sodium tetraborate networks.

On similar consideration, the 1·5 MEV $^{7}_{3}\mathrm{Li}$ and 1 MEV $^{4}_{2}\alpha$ particles resulting from slow neutron fissions of the boron cause approximately 4×10^{16} displacements per day, also in the sodium tetraborate networks.

Hence, the total number of displacements in sodium tetraborate is approximately $4{\cdot}4 \times 10^{16}$ per day.

Now figure 10 shows an initial pile bleaching rate of approximately 4×10^{16} centres per day, which suggests that for every atom displaced in the $\mathrm{Na_2B_4O_7}$ networks approximately one colour centre is destroyed.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to their colleagues, in particular Dr. H. M. Finniston, for their many helpful discussions concerning this report, and to Mrs. D. W. Jones for carrying out the readings on the Unicam Spectrophotometer.

REFERENCES

BEST, J. V. E., 1953, *The Coloration of some Optical Glasses* (London: H.M.S.O.).
BROWN, C. S., and THOMAS, L. A., 1952, *Nature, Lond.*, **169**, 35.
DAINTON, F. S., and ROWBOTTOM, J., 1954, *Trans. Faraday Soc.*, **50**, 488.
FRÖHLICH, H., 1947, *Proc. Roy. Soc.* A, **188**, 521.

Gale, A., and **Bickford**, F., 1953, *Nucleonics*, **11**, 48.
Mayer, G., and **Guéron**, J., 1952, *J. Chim. Phys.*, **49**, 204.
Mitchell, E. W. J., and **Paige**, E. G. S., 1954, *Proc. Phys. Soc.* B, **67**, 262.
Monk, G. S., 1952, *Nucleonics*, **10**, 52.
Mott, N. F., and **Gurney**, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford : University Press).
Pohl, R. W., 1937, *Proc. Phys. Soc.*, **49** (Extra Part), 3.
Seitz, F., 1946, *Rev. Mod. Phys.*, **18**, 384; 1949, *Disc. Faraday Soc.*, **5**, 271.
Simpson, J. H., 1950, *Proc. Phys. Soc.* A, **63**, 86.
Smakula, A., 1930, *Z. Phys.*, **59**, 603.
Sojka, O., 1940, *Die Verfärbung der Alkaliborate durch Radiumstrahlen* (Wien, Dissertation).
Thomas, L. A., *et al.*, 1949, *Disc. Faraday Soc.*, **5**, 341.
Turner, C. H. M., and **Lewis**, W. E., 1947, *Nature, Lond.*, **159**, 334.
Urbanek, J., 1951, *Acta Phys. Aust.*, **5**, 69.
Yokota, R., 1952 a, *J. Phys. Soc. Japan*, **7**, 316; 1952 b, *Ibid.*, **7**, 222; 1954, *Phys. Rev.*, **93**, 896.

Entrainment and Turbulence in Open Gas Streams

By R. H. GOUGH†
The Gas Research Board, Beckenham, Kent

MS. received 19*th July* 1954, *and in final form* 12*th November* 1954

Abstract. A method of studying the entrainment of air by a moving gas stream is presented which consists, in effect, of constructing entrainment curves for specified positions within the stream. Curves such as these have not, so far as is known, been published previously. They are of interest in showing the change from laminar to turbulent flow in the stream with particular clarity. A comparison with entrainment curves for gas burners is made.

The profile of the gas flow in the region of the laminar–turbulent transition region has also been studied experimentally, and a correlation obtained with the previous results.

§ 1. Introduction

In investigating the entrainment of primary air by a gas burner, it is usual to show the results in the form of an entrainment curve, in which the total air entrained by the jet is plotted against the gas flow (figure 6 (*b*)). The present method of interpreting these curves theoretically is to calculate the entrainment from the conservation of momentum principle and energy considerations (Silver 1948). The equation so obtained gives a fairly good agreement with experiment, and can be improved by taking into consideration extra factors such as gas buoyancy, although only at the expense of extra complexity in the final equation. The method is also only feasible for relatively simple systems, e.g. those having plain cylindrical mixture tubes.

In the present paper, the dependence of entrainment on the flow conditions in a free gas stream is shown by means of entrainment curves plotted for specific points within the stream. These provide a clear picture of the large and sudden change in entrainment between the laminar and turbulent regions of the stream, and also of the position of the break region between the two.

By comparing the curves with the usual type of gas burner entrainment curve, the characteristic rise in entrainment of the latter with increasing gas flow is seen to be capable in part of interpretation as a gradual change from laminar to turbulent conditions in the gas stream.

Proceeding further from the above, the profiles of flow in the laminar and turbulent conditions have been studied experimentally, and some suggestions put forward as to the possibility of deducing entrainment in gas burners from a physical picture of the flow conditions in the entraining section.

Two gas flow systems have been investigated. One is the straightforward case of a free jet in air. The other is that of the radial gas film formed by the impact of a gas jet on a plane surface, and it is actually on this system that the more comprehensive work has been carried out. The reason that such a system was chosen is that the work reported in this paper developed out of a previous

---

† Private address : 12 Hunston Road, Sale, Manchester.

programme which was concerned with investigating the mechanics of the target impact burner (Simmonds and Willett 1951), a system in which combustion takes place in a radial stream of fuel gas produced by jet impact in the manner stated above.

§ 2. METHOD AND APPARATUS

The object of the study was to measure the entrainment at a given position in the field of flow outside an orifice from which gas is issuing. This was to be done over a range of gas flows, covering the regions of both laminar and turbulent flow. The method adopted was to withdraw samples by means of a hypodermic needle whose tip was placed in the stream and then to analyse these for mixture composition. Carbon dioxide was chosen as the working fluid for the following reasons: (i) ease of analysis: a Haldane apparatus was used for measuring the CO_2 content of the samples; (ii) non-toxicity; (iii) high density: the jet (directed downwards) and sampling apparatus were placed within a box having openings to atmosphere at top and bottom, and it was assumed that all the CO_2 would eventually leave the box from the bottom.

Figure 1 shows the gas delivery tube (diameter 0·7 mm) and sampling arrangement for the radial flow measurements. The hypodermic needle can be moved by micrometer screws in the horizontal and vertical directions over the steel plate. For the experiments on a free jet the steel plate was removed and a vertical sampling needle used. The gas flow rate was measured by means of a variable-orifice flowmeter. The aerated gas sample was collected in a glass bulb from which mercury flowed out at a rate controlled by a capillary tube constriction. An appropriate sampling rate was calculated by ensuring that the gas velocity in the hypodermic needle was less than that in the external gas and air stream. After collection, the samples could be transferred directly to the Haldane apparatus and analysed for CO_2 content.

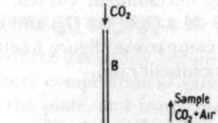


Figure 1. Apparatus for production and sampling of radial gas stream. A, polished steel plate; B, CO_2 delivery tube; C, hypodermic sampling needle movable in radial and vertical directions.

The degree of accuracy of the measurements with this apparatus is obviously governed by the fact that the gas jet must be some degree in contact with the outside atmosphere in order to ensure the disposal of used gas, and is therefore subject to the effects of draughts. These effects have been minimized by allowing the minimum contact between the jet and room air, and by ensuring a reasonably still atmosphere outside the box. The degree of reproducibility of results at the lower flow rates is within approximately 1% and gives a measure of the success of these precautions.

§ 3. Composition–Flow Curves for Radial Flow

The curves of figure 2 show the fraction of carbon dioxide in the sample plotted against gas flow in the nozzle for a given radial distance *a* of the sampling needle. In all the measurements the needle was resting upon the plate surface, and the outlet of the gas delivery tube was at a height of ½ cm above the plate.

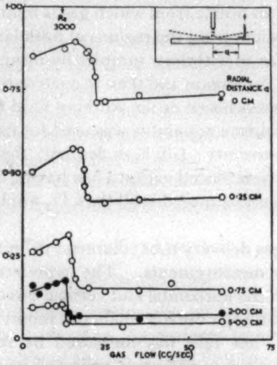


Figure 2. Composition–flow curves for a radial gas stream in air.

The curves are presented in terms of gas fraction in the mixture rather than air entrainment, since if the latter had been chosen it would have been difficult to show all the curves upon one graph. The curves are in shape approximately the inverse of the corresponding entrainment curves. The relationship between fraction of gas in the mixture $M=Q_g/(Q_a+Q_g)$ and entrainment ratio $R=Q_a/Q_g$ is simply $R=M^{-1}-1$. For comparison, figure 6 (*a*) shows a curve of the above type redrawn in terms of entrainment ratio.

§ 4. Evaluation

Upon inspection of the curves it will be noticed that they all correspond approximately in general outline, consisting of a slowly rising section at the lower range of flows, then a sharp drop in CO_2 content to a lower and almost constant value. This sharp drop, which corresponds to a sudden rise in entrainment, is seen to amount almost to a discontinuity in the curve, and might in fact prove to be so were it possible to sample accurately at a given point instead of over the area of the sampling needle (diameter 0·25 mm). This entrainment rise is apparently due to a change in flow conditions at the sampling point, i.e. from laminar flow at the lower gas flow rates to turbulent flow at the higher rates. This view is strengthened by noting that the break position occurs at lower flow rates for greater radial distances of the sampling needle. Thus turbulence commences at the outer regions of the stream and moves progressively inwards as the flow rate is increased, as is the case with a free jet in the air. In fact it is possible to compare the two very closely in the following manner. From the curves of figure 2 radial distance to position of break can be plotted against Reynolds number in the tube (figure 3). For comparison the right-hand side of the figure shows a typical 'height to turbulence' graph for an open jet, as determined by Scholefield and Garside (1949) by shadow photographs of unignited gas streams. It appears that the 'distance to turbulence' graph for the radial flow case is the equivalent of the section marked AB in the free jet 'height to turbulence' graph.

The entrainment curves for specific points in the field thus show very exactly (*a*) the change in entrainment between laminar and turbulent conditions of the flow, and (*b*) the position of the laminar–turbulent break for a given flow rate.

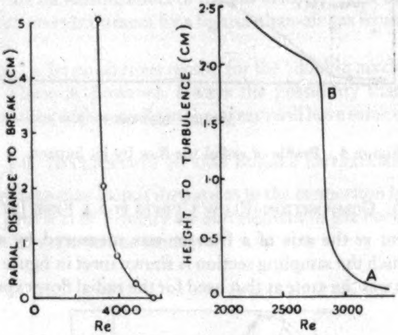


Figure 3. Left, variation of flow break position in radial gas flow. Right, a typical 'height to turbulence' curve for free jet.

§ 5. Flow Profile at the Break to Turbulence

In order to obtain a profile of the gas stream before and behind the break it was necessary to measure the mixture composition at various distances from the axis, and at various heights above the plate, and from this information to build up a picture of the flow. It will be realized that in the case of radial gas flow it is not possible to take shadow or Schlieren photographs through the stream as can be done with a free jet, since the disc of gas in front of and behind the section which it is desired to photograph will interfere excessively.

A gas flow of 16 cm^3 sec^{-1} was used in the apparatus already described. From figure 3, this gives a break position of 0·585 cm from the centre. At five positions about this break point, the gas composition was measured at various heights above the plate. From these measurements, it was possible to find the height at which the CO$_2$ content had decreased to an arbitrary standard value and to plot the contour of this height along the plate. The standard decided on was that of one half the maximum CO$_2$ fraction at each radial distance. The curve is shown in figure 4. It reveals a rapid thickening of the stream beginning at the break point, which appears to be more pronounced than that seen in free jets by means of shadow photographs such as those in Scholefield and Garside's paper. It also confirms the break position of 0·585 cm from the jet. It may be pointed out that a contour for a standard of 1/10 of the maximum CO_2 fraction showed the same general outline at a greater height above the plate surface.

The break shown by the flow profiles therefore correlates with that shown by the entrainment curves.

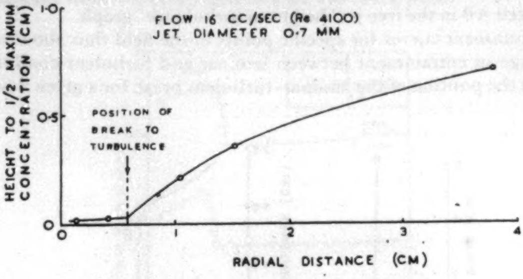


Figure 4. Profile of radial gas flow by jet impact.

§ 6. Composition–Flow Curves for a Free Jet

Air entrainment at the axis of a free jet was measured by an experimental arrangement of which the sampling section is shown inset in figure 5. Apart from this, the apparatus was the same as that used for the radial flow experiments. The sampling rate was enabled to be higher (~ 4 cm^3 min^{-1}) owing to the fact that the fall-off in velocity in a free jet is far less than in the radial stream formed by jet impact. As before, it was calculated by equating the gas velocity in the sampling needle to the minimum gas velocity encountered on the axis of the jet. The concentration–flow curves obtained at six different distances from the gas outlet are shown in figure 5. All are for axial positions of the sampling needle in the gas jet.

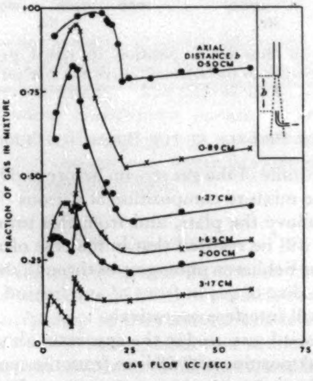


Figure 5. Composition–flow curves for a free gas jet in air.

§ 7. Evaluation

The top curve, for distance 0·5 cm, is identical with that for zero radial distance in the previous set. The other curves at increasing distances from the nozzle show a gradual smoothing-out of the sharp drop in CO_2 content, so that the latter occurs over a range of flows instead of at one particular flow. The reason for this may be the difficulty of aligning the jet and needle precisely. In the present case, alignment was carried out mechanically. A more satisfactory method would probably be to fix the axis of the gas jet by determining the position of maximum gas pressure in the stream cross section.

The appearance of a sharp rise before the drop in the lower curves is almost certainly due to the increasing effect of the gas density. It is the equivalent of the effect of buoyancy on entrainment by a lighter-than-air gas issuing from an upward pointing nozzle.

There appears to be no obvious reason for the ' double maximum ' on the lower three curves. There is, however, always the possibility that with this type of sampling the presence of the needle in the stream will have some effect on the results.

§ 8. Application to Gas Burner Entrainment

A few suggestions may be put forward as to the connection between the present curves and the problem of primary air entrainment in gas burners, though at this stage they must be regarded as tentative.

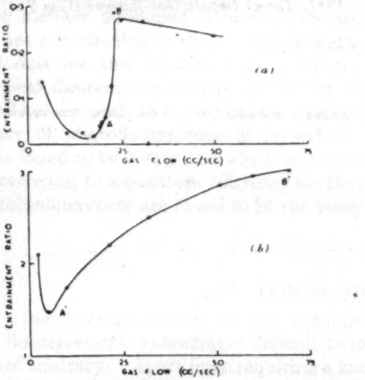


Figure 6. (*a*) Entrainment curve at the axis of a free gas jet. (*b*) Typical gas burner entrainment curve.

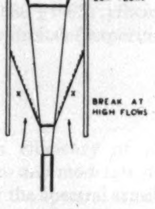


Figure 7. Flow conditions in the entrainment section of a gas burner. X=Turbulent region.

In figure 6 (*a*) is shown an entrainment curve for a point on the axis of a free jet, calculated from the upper curve of figure 5. For comparison, figure 6 (*b*) shows a typical entrainment curve for primary air entrainment in a Bunsen type burner. Since the sharp rise AB in (*a*) is due to the onset of turbulence in the stream, it is reasonable to regard the portion A′B′ of (*b*) as being due in part to the same cause. It appears that the more gradual rise in entrainment in the gas burner is due to the fact that we are considering the whole section of the flow bounded by the burner tube wall instead of at one point.

The profile graph of figure 4 suggests a possible method for predicting the entrainment to be expected in a given burner. Figure 7 shows the equivalent profile in a burner tube, assuming that the back pressure effect of the burner ports is not so great as to affect the stream profile appreciably. Such a diagram could be drawn to scale knowing the height to turbulence for the given flow, and the divergence angles of the laminar and turbulent sections. There is evidence to show that these angles remain constant with flow (Tollmien 1926, Farquharson 1952). From the diagram, the height above the nozzle at which the cross section of the gas stream meets the burner wall can be found, and above this point no further entrainment will take place. If this height is known the resulting entrainment could then be read off at any given gas flow from a corresponding entrainment curve.

It is hoped to test this method of predicting entrainment as soon as the necessary practical data has been obtained.

REFERENCES

FARQUHARSON, I. M. C., 1952, *J. Instn Heat. Vent. Engrs*, **19** (197), 449.

SCHOLEFIELD, D. A., and GARSIDE, J. E., 1949, *The Structure and Stability of Diffusion Flames* (Gas Res. Board Pub. GRB 48).

SILVER, R. S., 1948, *The Calculation of Air Entrainment in Gas Burners* (Gas Res. Board Pub. GRB 39/23).

SIMMONDS, W. A., and WILLETT, A. L., 1951, *Target Impact Gas Burners* (Gas Res. Board Pub. GRB 65).

TOLLMIEN, W., 1926, *Z. angew. Math. Mech.*, **6**, 468.

Absolute Quantum Efficiency of Photofluorescence of Anthracene Crystals

By G. T. WRIGHT
Department of Physics, Rhodes University, Grahamstown, South Africa

MS. received 12*th October* 1954

Abstract. The fluorescence characteristics of organic crystals differ appreciably from those of the emitting molecules due to self-absorption of the molecular fluorescence. These differences are investigated and properties of the fluorescence of the crystal are related to the corresponding properties of the molecular fluorescence; the equations developed are made the basis of an experimental determination of molecular quantum efficiencies of photofluorescence for the crystalline state. This method avoids the considerable errors involved in a direct measurement which are occasioned by having to integrate numerically over the whole sphere of emission inside an optically anisotropic crystal the fluorescence intensity that is measured directly only within a certain solid angle external to the crystal.

The method developed is particularly applicable to crystalline anthracene which has considerable overlap of absorption and fluorescence spectra. Experimental data for this substance are presented. These include the molecular and crystal fluorescence spectra and decay times; the latter are found to be $6{\cdot}4 \pm 0{\cdot}2$ mμsec and $18{\cdot}0 \pm 0{\cdot}5$ mμsec respectively. The molecular quantum efficiency of photofluorescence of crystalline anthracene at a temperature of 290°K is found to be $0{\cdot}94 \pm 0{\cdot}02$ which is reduced, by the effects of self-absorption of fluorescence, to a quantum efficiency for the crystal of $0{\cdot}80 \pm 0{\cdot}05$; efficiencies of photoluminescence are found to be the same within the limits of experimental error.

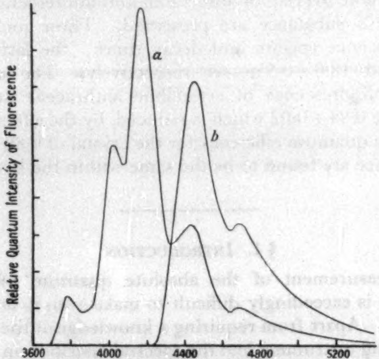
§ 1. Introduction

A direct measurement of the absolute quantum efficiency of photofluorescence is exceedingly difficult to make even to any moderate degree of accuracy. Apart from requiring a knowledge of the spectral sensitivity curve of the detecting instrument and the spectral distribution of the fluorescence emission, the fluorescence intensity is measured directly only within a certain solid angle and this has to be integrated over the whole sphere of emission; assumptions made for this purpose are never very satisfactory and are the major source of error in a measurement of this nature (see for instance Gilmore *et al.* 1952, Forster and Livingstone 1952). Even for liquids and transparent glasses the problem of ascertaining the precise fraction of the fluorescence received is complicated by the effects of anisotropic emission of fluorescence (Ganguly and Chaudhury 1954) and re-absorption of fluorescence; for the crystalline state the effects of optical anisotropy constitute an almost insurmountable obstacle in a direct determination. For this reason reliable measurements of quantum efficiencies for the crystalline state are completely lacking although such measurements are of fundamental importance. In this communication an indirect method is described for the experimental determination of quantum efficiencies for crystals.

It is well established that the photofluorescence of organic materials is inherently a molecular property which nevertheless can be considerably modified by the physical environment of the molecule. In addition, however, re-absorption of fluorescence which occurs when the absorption and fluorescence spectra of the molecule overlap modifies the characteristics of the molecular fluorescence. In the present communication the decay time and quantum efficiency of the fluorescence of the crystal are related to the corresponding molecular quantities from which they differ due to the effects of spectral overlap and the equations developed used as the basis of an experimental determination of photofluorescence quantum efficiencies for the crystalline state.

§ 2. CRYSTAL FLUORESCENCE

The most obvious modification of the molecular emission is suppression of a part since only those wavelengths to which the crystal is transparent are observed. This has been studied in detail for some crystals by Wright (1954) and Birks and Wright (1954) and the existence of very considerable overlap verified in certain cases, notably anthracene. The molecular and crystal fluorescence spectra of this material are reproduced in the figure to illustrate the extent of the effects of spectral overlap.



Fluorescence spectra of crystalline anthracene at 290°K : *a*, fluorescence of microcrystals; *b*, fluorescence by transmission through 1 cm crystal. The areas of these curves are adjusted to be proportional to the molecular and crystal quantum efficiencies respectively.

In addition, the decay time of the crystal fluorescence is longer, due to the re-absorption processes occurring in the overlap region (Birks and Little 1953) and the quantum efficiency is less. For excited molecules surrounded by a large number of unexcited molecules these effects may be readily calculated.

Each excited molecule has probability $p\,dt$ in time dt of losing its excitation energy radiatively and probability $k\,dt$ of losing it non-radiatively. The rate of disappearance of excited molecules is accordingly $p+k$ and the quantum efficiency is $q=p/(p+k)$. However, since the crystal is opaque to part of the molecular emission each photon has only chance κ of being emitted in that part of the fluorescence spectrum to which the crystal is transparent and thus of escaping through the crystal. Consequently, the probability of emission from the crystal is reduced to $p\kappa\,dt$. For the crystal then the rate of disappearance of excited molecules is $p\kappa+k$ and the quantum efficiency is

$$q'=p\kappa/(p\kappa+k). \qquad \ldots\ldots(1)$$

In general, fluorescence is excited with wavelengths for which the extinction coefficient of the crystal is high compared with its value in the region of spectral overlap. This results effectively in surface excitation of the crystal and a more involved discussion is necessary. For initial simplicity we shall consider a semi-infinite crystal.

(i) *Quantum Efficiency of Fluorescence of Crystal*

If the total number of excited molecules initially is N_0 a number qN_0 emit fluorescence photons of which a fraction δ, determined by the optical constants of the crystal, escape immediately through the surface of the crystal. The remainder of the photons, $q(1-\delta)N_0$, travel into the crystal, a proportion κ finding it transparent and a proportion $1-\kappa$ being re-absorbed to produce further excited molecules. The photons emitted by these molecules follow the same sequence of events and the whole cascade process continues until eventually no excited molecules remain.

The total number of photons emitted through the surface is:

$$N_s=q\delta N_0[1+(1-\delta)(1-\kappa)q+(1-\delta)^2(1-\kappa)^2q^2+\ldots.]$$

$$N_s=\frac{q\delta N_0}{1-q(1-\delta)(1-\kappa)} \qquad \ldots\ldots(2)$$

and the number escaping into the crystal is:

$$N_c=q\kappa(1-\delta)N_0[1+(1-\delta)(1-\kappa)q+(1-\delta)^2(1-\kappa)^2q^2+\ldots.]$$

$$N_c=\frac{q\kappa(1-\delta)N_0}{1-q(1-\delta)(1-\kappa)}. \qquad \ldots\ldots(3)$$

The ratio $(N_s+N_c)/N_0$ gives the quantum efficiency of the crystal fluorescence as:

$$Q=\frac{q(\delta+\kappa-\delta\kappa)}{(1-q)+q(\delta+\kappa-\delta\kappa)}. \qquad \ldots\ldots(4)$$

When the escape coefficient κ is unity this reduces to $Q=q$ but when κ is less than unity Q is less than q.

(ii) *Decay Time of Crystal Fluorescence*

The rate of disappearance of excited moleules is:

$$dN=-\{k+p[1-(1-\delta)(1-\kappa)]\}N\,dt=-\lambda N\,dt \qquad \ldots\ldots(5)$$

giving $N=N_0e^{-\lambda t}$.

The intensity of the emission through the surface is $I_s = p\delta N dt$ and the intensity of that escaping through the crystal is $I_c = p\kappa(1-\delta)N dt$. We have:

$$N_s = \int_0^\infty I_s = \frac{p\delta N_0}{k+p(\delta+\kappa-\delta\kappa)} \qquad \ldots\ldots(6)$$

$$N_c = \int_0^\infty I_c = \frac{p\kappa(1-\delta)N_0}{k+p(\delta+\kappa-\delta\kappa)} \qquad \ldots\ldots(7)$$

in agreement with equations (2) and (3) for N_s and N_c respectively when the substitution $q = p/(p+k)$ is made.

The decay time of the crystal emission is obtained from equation (5):

$$\tau_c = [k+p(\delta+\kappa-\delta\kappa)]^{-1} \qquad \ldots\ldots(8)$$

which may be compared with the decay time of the molecular emission:

$$\tau_m = (p+k)^{-1}. \qquad \ldots\ldots(9)$$

The only unknown quantity in equations (1) to (9) is the molecular quantum efficiency q since in principle both κ and δ may be evaluated from optical data and both τ_c and τ_m can be measured. Solving equations (8) and (9) for the molecular quantum efficiency we have:

$$q = \frac{1-\tau_m/\tau_c}{1-(\delta+\kappa-\delta\kappa)}. \qquad \ldots\ldots(10)$$

This equation for q contains quantities which can be readily determined and therefore forms a suitable basis for an experimental measurement of the quantum efficiency; its application to the particular case of anthracene is described in the next section. Anthracene was chosen since for this material the quantities τ_m/τ_c and $\delta+\kappa-\delta\kappa$ are both much smaller than unity; this makes the result for q relatively insensitive to experimental errors of measurement.

§ 3. EXPERIMENTAL

The fluorescence characteristics of the crystal were measured using the experimental approximation of a 1 cm cube crystal. This is justifiable since these dimensions are large compared with mean photon path lengths for re-absorption which mostly range between 10^{-4} cm and 10^{-1} cm in the region of spectral overlap. Anthracene crystals as supplied by the Larco Nuclear Instrument Company for scintillation counting were used, both transmission and reflection spectra being observed.

For observation of the molecular fluorescence characteristics microcrystalline layers were used in which self-absorption was negligible. These were prepared by allowing a few drops of a weak solution of anthracene to evaporate on a clean glass disc. In this way thin, uniform layers were obtained.

The fluorescence spectra were measured using a Cenco grating monochromator with a photomultiplier detector. Calibration curves for these instruments had been previously measured and allowed the spectra to be plotted in terms of relative quantum intensities.

(i) *The Absorption Escape Coefficient* κ

This quantity is readily determined from a comparison of the molecular fluorescence spectrum, obtained from the microcrystalline layers, and the spectrum given by transmission through a 1 cm crystal. These are the spectra shown in the figure. In the region to which the crystal is completely transparent these spectra are identical as expected and if normalized at these wavelengths the ratio of their areas gives the absorption escape coefficient κ. This procedure is valid since the spectra are plotted in terms of relative quantum intensities. For anthracene we find $\kappa = 0{\cdot}24$.

(ii) *The Optical Escape Coefficient* δ

Anthracene is optically very anisotropic having principal refractive indices for the crystal of 1·959, 1·786 and 1·555. However, the crystal was considered isotropic with a mean index of 1·767 and δ calculated as 0·089; the effects of optical anisotropy and re-absorption of fluorescence are discussed later.

(iii) *Measurement of Decay Times*

These have been previously measured for anthracene by other workers. However, the results obtained using various techniques are not in very close agreement and accordingly these measurements were repeated. The decay times were determined by measuring the relative modulation of the fluorescence light and the exciting light using an apparatus of improved design which is to be described elsewhere. The exciting light covered the wavelength range 3000 Å to 3900 Å and could be modulated at a frequency of 25·80 Mc/s per second or 38·70 Mc/s per second. Measurements at each frequency gave the same results: $\tau_{\mathrm{m}} = 6{\cdot}4 \pm 0{\cdot}2$ mμ sec, $\tau_{\mathrm{c}} = 18{\cdot}0 \pm 0{\cdot}5$ mμ sec.

(iv) *Calculation of the Quantum Efficiency* q

Before these measurements are substituted in equation (10) to obtain the molecular quantum efficiency of fluorescence it is necessary to consider the various sources of error inherent in this determination.

The most serious is due to uncertainty in the estimated value of the optical escape coefficient δ. Since all organic crystals are optically anisotropic δ has a different value for different crystal surfaces. This is shown experimentally by small variations of the decay time of the fluorescence when different crystal faces are excited. The differences observed indicate variations in δ of about 15%.

The value of δ is also modified by photon migration in the cascade process of re-absorption and emission. After the initial emission, re-absorption of photons produces excited molecules lying deeper within the crystal. Even though these may then emit photons towards the surface a large number of these will be re-absorbed before they can reach the surface and escape. This reduces the effective value of δ. The magnitude of this reduction can be obtained by comparing the molecular spectrum with the reflection spectrum from the crystal surface. When these are normalized at long wavelengths they should be identical but re-absorption of photons emitted towards the surface causes the short wavelength peaks of the reflection spectrum to be depressed (Birks and Wright 1954). The ratio of the areas of these spectra is 0·46 and as a first approximation we may consider that δ is reduced in this ratio.

A further modification of the value of δ is caused by surface irregularities. Because of their softness anthracene crystals cannot be optically polished and in any event constant sublimation produces surfaces with a slightly translucent appearance. These surface irregularities increase the effective value of δ and thus oppose the decrease caused by photon migration. The resultant effective value of δ is obtained most readily by experimental methods.

It is evident from equations (2) and (3) that the fraction of fluorescence received by a detecting instrument looking at reflected or transmitted fluorescence is sensitive to the magnitude of δ whereas the quantum efficiency is little affected. This may be used to obtain an estimate of δ. The first absorption maximum of crystalline anthracene occurs at $\lambda = 3950$ Å; excitation in this wavelength region therefore corresponds to surface excitation for which equations (2) and (3) are applicable. The transmission limit of the crystal was measured to be at $\lambda = 4240$ Å; excitation with light of slightly shorter wavelength thus produces excited molecules lying deep within the crystal and equation (1) is applicable. Freshly cleaved surfaces were polished smooth on soft tissue and used for these measurements, being excited with monochromatic light obtained from a Beckman Model D.U. absorption spectrophotometer used as a monochromator. At each wavelength the intensity of the exciting radiation was normalized to the same quantum flux using the calibrated photomultiplier. As expected the intensity of the transmitted fluoresence was greater when excited with the longer wavelengths, the mean ratio of intensity for several measurements being 1·25 : 1. Using the calculated value $\delta = 0{\cdot}089$ in equation (10) an approximate value for the quantum efficiency was estimated and used in equations (1) and (3), from which an effective value of $\delta = 0{\cdot}10$ was obtained. The accuracy of this measurement is estimated at $\pm 10\%$.

Thus far it has been tacitly assumed that for a given material κ is a constant determined only by the extent of overlap of the molecular absorption and fluorescence spectra. Since the absorption edge of the crystal is not vertical, however, the value of κ will depend slightly upon the dimensions of the crystal as a very small fraction of the emitted photons will have mean path lengths for re-absorption of the order of the crystal dimensions. To reduce this source of error all measurements were made with the same size of crystal in order that the value of κ, deduced from the spectral curves, and the decay time measurements could be used together. The uncertainty in κ is thus due predominantly to experimental error and a generous estimate of $\pm 5\%$ will be allowed for this.

Lastly the assumption that use of microcrystalline layers enables the true molecular fluorescence to be observed requires some justification. The estimated thickness of the layers used varied between $0{\cdot}5\,\mu$ and $0{\cdot}05\,\mu$ and while it is certain that there must be a small amount of re-absorption taking place very little difference was observed in the spectrum and decay time of fluoresence from different layers. We may therefore conclude that re-absorption is very small in layers of these thicknesses.

Substituting the above experimental results in equation (10) the photofluorescence quantum efficiency of the anthracene molecule in the crystalline state and at a temperature of 290° K, is found to be $q = 0{\cdot}94 \pm 0{\cdot}02$. The quantum efficiency of the surface excited crystal fluorescence, which is that generally observed, is given by equation (4) as $0{\cdot}83 \pm 0{\cdot}05$ while for the infinite crystal equation (1) gives a quantum efficiency of $0{\cdot}80 \pm 0{\cdot}05$.

It should be noted that the phosphorescence of anthracene is not involved in the present measurement; this is due to transitions from metastable triplet levels and will be considered now. The direct radiative transition from the first triplet state of the molecule to the ground state gives an exceedingly weak emission at longer wavelengths than the fluorescence emission and may be neglected (Kasha 1944, Reid 1952). The corresponding absorption band is too weak to be detected. Accordingly, for illumination in the fundamental absorption band, the triplet state is reached by a non-radiative transition from the first excited singlet state. The formation of excited triplet states thus constitutes a mechanism for non-radiative depopulation of the excited singlet states. In the previous discussion this process is thus included in the non-radiative transition constant k. Thermal de-activation of the triplet states back into the singlet states gives phosphorescence with the same spectrum as fluorescence and results in the luminescence efficiency being greater than the fluorescence efficiency. However, since the phosphorescence/fluorescence intensity ratio for crystalline anthracene is so very small that it may be neglected, we may conclude that, within the limits of experimental error, the photoluminescence quantum efficiency is the same as the photofluorescence quantum efficiency. This conclusion is supported by the measurements of Bowen *et al.* (1949) who have observed that the luminescence efficiency of powdered anthracene increases only slightly at liquid air temperature and from this and other relevant data infer that the quantum efficiency at this temperature is unity. This gives a photoluminescence quantum efficiency of 0·9 at room temperature agreeing with the present measurement.

§ 4. CONCLUSION

The experimental method described for the measurement of photofluorescence quantum efficiencies for crystals is particularly suitable for substances such as anthracene which have a large overlap of absorption and fluorescence spectra. The uncertainties which are introduced into a direct determination by the very anisotropic optical properties of the crystal are largely eliminated with the present technique and reliable results are given notwithstanding considerable uncertainty in measurement of the optical data.

For compounds which cannot be obtained in suitable crystalline form or which do not possess sufficient spectral overlap an indirect method may be used based on comparison with the established value for anthracene. Direct comparison of relative fluorescence intensities is not reliable even if the spectral response of the detecting instrument is known for equations (2) and (3) show that although the optical escape coefficient δ hardly affects the result for the quantum efficiency it strongly influences the proportion of fluorescence received by a detecting instrument looking at reflected or transmitted fluorescence. A more reliable method is to use the area under the absorption curve as a measure of the radiative transition probability p, from the first excited level and the microcrystalline decay time, τ_{m}, as a measure of the total transition probability. Using the equation $q=p\tau_{\mathrm{m}}$ the quantum efficiency may be found by comparison with the corresponding values for anthracene.

ACKNOWLEDGMENTS

The author is indebted to Dr. M. E. Szendrei for providing the calibration curves for the monochromator and photomultiplier used and to Mr. T. D. S. Hamilton for making the decay time measurements. Thanks are also due to the South African Council for Scientific and Industrial Research for a grant to purchase equipment.

References

Birks, J. B., and Little, W. A., 1953, *Proc. Phys. Soc.* A, **66,** 921.

Birks, J. B., and Wright, G. T., 1954, *Proc. Phys. Soc.* B, **67,** 657.

Bowen, E. J., Mikiewicz, E., and Smith, F. W., 1949, *Proc. Phys. Soc.* A, **62,** 26 (and private communication).

Forster, L. S., and Livingstone, R., 1952, *J. Chem. Phys.*, **20,** 1315.

Ganguly, S. C., and Chaudhury, N. K., 1954, *Phys. Rev.*, **95,** 1148.

Gilmore, E. H., Gibson, G. E., and McClure, D. S., 1952, *J. Chem. Phys.*, **20,** 829.

Kasha, M., 1944, *J. Amer. Chem. Soc.*, **66,** 2100.

Reid, C., 1952, *J. Chem. Phys.*, **20,** 1214.

Wright, G. T., 1954, Paper read before S.A. Congress of Physics, Pretoria, July.

Thermoluminescence of Alkali Halide Phosphors

By H. N. BOSE†

Khaira Laboratory of Physics, University College of Science, Calcutta, India

MS. received 2nd February 1954 and in amended form 25th November 1954

Abstract. Thermoluminescence curves of NaCl, KCl, NaCl : Tl, KCl : Tl and mechanical mixtures of TlCl with NaCl and KCl have been recorded. The samples have been excited by 10 kv cathode rays at liquid oxygen temperature, and a high rate of heating (5–6 deg $\sec^{-1}$) has been employed. The emissions and the temperatures have been continuously recorded during heating. The trap depths of these phosphors have been estimated and the effects of heat treatment on the samples have been studied.

§ 1. Introduction

When an alkali halide is excited at low temperature, energy is stored in the various trapping centres, and it is indicated by the long period phosphorescence. The manner in which energy is released at different temperatures, when the temperature of the sample is being steadily raised, indicates the nature of traps present in the phosphor. The shallow traps give the luminescence peaks at low temperatures and the deeper ones are released at higher temperatures. Thus thermoluminescence provides us with a method of investigating the trap depths in a phosphor.

Randall and Wilkins (1945), Garlick and Gibson (1947), Bull and Garlick (1950) have developed approximate methods of determining the trap depth from the thermoluminescence curve. The frequency factor s for alkali halide phosphors has been assumed by some of the authors (Dutton and Maurer 1953) to be of the order of $10^{9\pm1}\ \sec^{-1}$; the approximate formula (Randall and Wilkins 1945, Leverenz 1950) for E, the trap depth in ev, if this value of s is used, reduces to $E = 21\boldsymbol{k}T_{\mathrm{G}}$ where $\boldsymbol{k}$ is the Boltzmann's constant and T_{G} the absolute temperature of maximum glow.

In the present work, measurements on NaCl, NaCl : Tl, 2·5%, KCl, KCl : Tl, 2·5% have been reported. The samples were pure reagents of E. Merck and the same as were used in the study of fluorescence spectra (Bose and Sharma 1950) and decay of the afterglow (Bose and Sharma 1953). The activated samples were prepared by mixing requisite amounts of thallium chloride and melting the mixture in an electric furnace.

Although a low rate of heating (0·25 deg $\sec^{-1}$) is usually preferred for the study of thermoluminescence, yet it has been observed that in the case of these phosphors, the weaker peaks get flattened out unless the rate of heating is high. So throughout the present work a high rate of heating was maintained. A fine film of powdered phosphor on the sample holder has been found more suitable for this purpose than a single crystal.

---

† Now Assistant Professor of Physics, Indian Institute of Technology, Kharagpur, India.

§ 2. EXPERIMENTAL TECHNIQUE

The demountable cathode-ray tube described previously (Bose and Sharma 1953) has been used in the present investigation. A thin-walled silver specimen holder was fitted with a copper–constantan thermocouple which, connected to a Moll type galvanometer ($T=1\cdot8$ sec), recorded temperature during heating. The experiment was carried out in a dark room and the photomultiplier, placed in front of the sample holder, was connected to a separate Moll type galvanometer ($T=1\cdot5$ sec). The photomultiplier tube was operated from a conventional type of electron tube regulated power unit, furnishing variable voltages between 800 and 1200 volts. The deflections of the two galvanometers were recorded photographically on a rotating drum. As fluorescence spectra of the phosphors were known (Bose and Sharma 1950), suitable photomultipliers (931A, 1P28, 1P22) could be selected for recording the thermoluminescence. The sample at liquid oxygen temperature was excited to saturation, with 10 kv cathode rays for a few minutes. The rate of heating employed was 5 to 6 deg sec^{-1}.

For recording the spectral nature of the glow peaks, two photomultipliers, sensitive in different regions, were made to record the thermoluminescence simultaneously during the same experiment, and optical filters were used to separate the emissions. Simultaneous records eliminated the effects of uncertainty due to individual excitation. While measuring the temperature of maximum glow from the records, necessary correction of the effects of the lagging of the galvanometer was made.

§ 3. RESULTS AND DISCUSSIONS

A summary of the results is given in the table. The thermoluminescence curve of pure NaCl is shown in figure 1, curve *a*. The peak emitted at 240°K is richer than the peaks at 650°K in ultra-violet emissions, while the latter is richer in emissions of the blue region; the low temperature peak at 165°K consists of visible as well as ultra-violet emissions. Introduction of thallium in NaCl has very little effect on the peak at 240°K and practically no effect on the peak at 165°K, but in place of the peak at 650°K, it produces two peaks at 375°K and 420°K (figure 2, curve *a*, the last peak has not been shown). The thermoluminescence of a mechanical mixture of NaCl and TlCl has been found similar to that of NaCl. The peaks at 225°K and 375°K of NaCl : Tl are comparatively rich in ultra-violet emissions (figure 2, curves *a* and *b*).

Sample	T_G (°K)	E (ev)†	Sample	T_G (°K)	E (ev)†
NaCl	(i) 165	0·30	KCl	(i) 120	0·22
	(ii) 240	0·44		(ii) 205	0·37
	(iii) ~650	1·17		(iii) 285	0·51
				(iv) ~620	1·12
NaCl : Tl 2·5	(i) 165	0·30	KCl : Tl 2·5	(i) 185	0·33
	(ii) 225	0·41		(ii) 230	0·42
	(iii) 375	0·68		(iii) 305	0·55
	(iv) 420	0·76		(iv) 390	0·71

† Based on assumption $s=10^{9\pm1}$ sec^{-1}.

The thermoluminescence of pure potassium chloride (figure 1, curve *b*) consists of a strong peak at 205°K and a few weaker peaks. The strong peak of KCl yields emissions, both in the visible and near ultra-violet region of the spectrum. Thallium activated KCl gives thermoluminescence (figure 3, curve *a*),

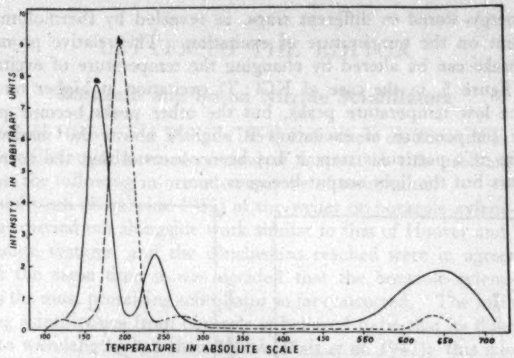


Figure 1. Thermoluminescence curves of *a* NaCl; *b* KCl, with RCA 931A photomultiplier tube.

consisting of four strong peaks, the peaks at 305°K and 390°K being comparatively rich in ultra-violet emissions (figure 4). The mechanical mixture of KCl and TlCl gives thermoluminescence somewhat similar to that of thallium activated KCl (figure 3, curve *b*).

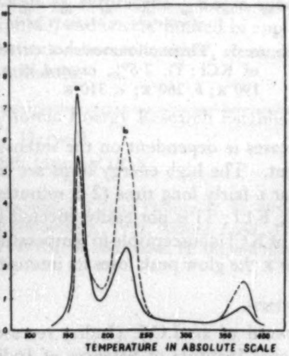


Figure 2. Simultaneously recorded thermoluminescence curves of NaCl : Tl, 2·5%, with *a*, 931A; *b*, 1P28.

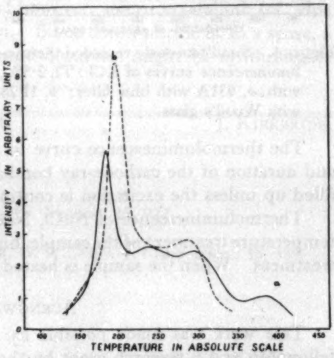


Figure 3. Thermoluminescence curves of *a*, KCl : Tl, 2·5%; *b*, mechanical mixture of KCl and TlCl, recorded with 931A.

The fluorescence spectra of alkali halides lie in the visible and ultra-violet portion of the spectrum (Bose and Sharma 1950). But, as shown in figures 2 and 4, and mentioned above, the thermoluminescence peaks differ in the spectral nature of their emissions. This leads to the conclusion that electrons from different traps, released at different temperatures, undergo preferential transitions during emission, so that the spectral natures of the peaks are not identical. This conclusion is also supported by the fact that the decay rate of phosphorescence varies in different parts of the spectrum (Bose and Sharma 1953).

The energy stored in different traps, as revealed by thermoluminescence, is dependent on the temperature of excitation. The relative prominence of different peaks can be altered by changing the temperature of excitation. As shown in figure 5, in the case of KCl : Tl excitation at higher temperatures reduces the low temperature peaks, but the other peaks become prominent. When the temperature of excitation is slightly above the maximum glow temperature of a particular trap, it has been observed that the corresponding peak appears but the light output becomes poor.

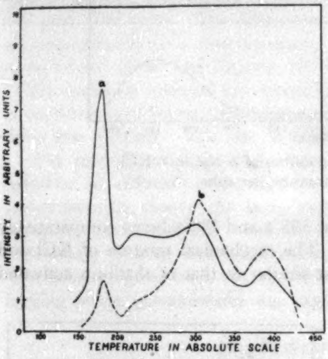


Figure 4. Simultaneously recorded thermoluminescence curves of KCl : Tl, 2·5%, with *a*, 931A with blue filter; *b*, 1P28 with Wood's glass.

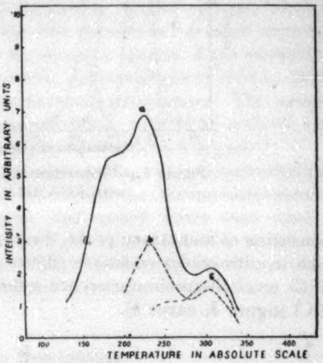


Figure 5. Thermoluminescence curves of KCl : Tl, 2·5%, excited at *a*, 190°K; *b*, 260°K; *c*, 310°K.

The thermoluminescence curve in most cases is dependent on the intensity and duration of the cathode-ray bombardment. The high energy traps are not filled up unless the excitation is continued for a fairly long time (2–5 minutes).

Thermoluminescence of NaCl, NaCl : Tl, KCl : Tl is not easily affected by temperature treatment of the sample, but that of KCl is susceptible to temperature treatment. When the sample is heated to 600°K the glow peak loses its intensity.

ACKNOWLEDGMENTS

This work was made possible by the award of an I.C.I. (India) research fellowship and a research grant by the National Institute of Sciences of India. The author is thankful to Professor S. N. Bose, Khaira Professor of Physics, Calcutta University, for his keen interest and helpful discussions. The author is also indebted to his student, Mr. J. Sharma, for help received during the experimental work.

REFERENCES

BOSE, H. N., and SHARMA, J., 1950, *Proc. Nat. Inst. Sci. India*, **26**, 47; 1953, *Proc. Phys. Soc.* B, **66**, 371.
BULL, C., and GARLICK, G. F. J., 1950, *Proc. Phys. Soc.* A, **63**, 1283.
DUTTON, D., and MAURER, R., 1953, *Phys. Rev.*, **90**, 126.
GARLICK, G. F. J., and GIBSON, A. F., 1947, *Proc. Roy. Soc.* A, **188**, 485.
LEVERENZ, H. W., 1950, *An Introduction to Luminescence of Solids* (New York : J. Wiley), p. 177.
RANDALL, J. T., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc.* A, **184**, 365.

LETTER TO THE EDITOR

Borazole and Boron Nitride Scintillators

The work of Hoover and Dohne (1954) and others on borazole as a neutron sensitive scintillator indicates a renewal of interest in this problem and it is thought that the following information may be of some value.

Published work (Kirkbride 1953) of the writer on borazole-xylene-terphenyl systems was carried out alongside work similar to that of Hoover and Dohne on borazole-solute systems, and the conclusions reached were in agreement with theirs. At the same time it was decided that the borazole-xylene-terphenyl system was the most promising scintillator so far concocted. The main difficulty in obtaining scintillations from borazole is believed to be that its fluorescence is restricted to wavelengths less than 2100 Å (Platt *et al.* 1947); this is some 600 Å less than exciton wavelengths readily transferred to solutes such as terphenyl or carbazole from aromatics such as xylene. It may be possible that a substituted borazole, e.g. phenyl borazole, may act as a solute which could bridge the gap between 2100 Å and the visible or near ultra-violet spectrum. It may also be possible that a substituted borazole may be a natural scintillator, for it was found, along with the original investigations (Kirkbride 1953), that well crystallized boron nitride was an excellent scintillator; it responded to alpha-particles at least as well as anthracene. Boron nitride is, however, micro crystalline (cf. zinc sulphide) and is very limited in application. Since boron nitride is, in a sense, a substituted borazole, other, clear, substituted borazoles might be synthesizable scintillators.

Atomic Energy Research Establishment,
Harwell.
8th January 1955.

J. Kirkbride.

Hoover, J. I., and Dohne, C. F., 1954, *Rev. Sci. Instrum.*, **25**, 922.
Kirkbride, J., 1953, *Nature, Lond.*, **171**, 564.
Platt, J. R., Klevens, H. B., and Schaeffer, G. W., 1947, *J. Chem. Phys.*, **15**, 598.

CORRIGENDUM

The Computation of Electric Dipole Moments, by E. A. Guggenheim (*Proc. Phys. Soc.* B, 1955, **68**, 186).

The left-hand side of the last (unnumbered) equation should read $Q(d_1/M_2)$.

REVIEWS OF BOOKS

Theoretical Elasticity, by A. E. Green and W. Zerna. Pp. xiv + 442. (Oxford : Clarendon Press, 1954.) 50*s*.

In this book three topics in the theory of elasticity are described in some detail, namely, finite strain, two-dimensional infinitesimal theory and the theory of shells, much of the material appearing here for the first time in text-book form. The development throughout rests entirely on general tensor methods, thus leading to a concise formulation of the theory.

The first section provides an introduction to tensor analysis, differential geometry of surfaces and some fairly sophisticated complex variable theory, followed by the formulation of finite strain theory. Stress and strain are defined and the compatibility relations follow naturally from the vanishing of the relevant Riemann–Christoffel tensor. The strain–energy function is then discussed and leads to the stress–strain relations. A few special problems are worked out in detail, in all of which a displacement is assumed and the system of forces responsible for the displacement is derived. The number of problems soluble in this way is very limited and they are rather artificial, but it is difficult to see how any other method can be adopted, since the governing equations are non-linear, the final shape of the deformed body is not known and the compatibility equations are of alarming complexity. Finally, there is a discussion of a small deformation superimposed on an initial finite one.

In the next section infinitesimal theory is derived as the limit of finite theory. Only a few general problems are considered (mainly indentation and crack problems), the main object being the discussion of two-dimensional problems using complex variable methods and an extension of these methods to non-isotropic media. A large class of problems can be treated in this way although most of the solutions have been found previously by other means.

The last section of the book is devoted to the general theory of shells, followed by a general discussion of various simple types of shell.

The book will be of great value to mathematicians who are already familiar with the physical aspects of elasticity (which are very much obscured by the formal treatment and concise notation), and who can cope with the tensor manipulation called for. Those familiar only with cartesian tensors will find the subject-matter very heavy going.

A. N. GORDON.

Rocket Exploration of the Upper Atmosphere, edited by R. L. F. Boyd and M. J. Seaton. Pp. viii + 376. (London : Pergamon Press, 1954.) 75*s*.

This is a collection of papers read at a conference arranged by the Upper Atmosphere Rocket Research Panel of the United States and the Gassiot Committee of the Royal Society at Oxford in August 1953. Work in this field has been carried out almost exclusively by American scientists who seized upon the opportunities offered by the capture of the German V-2 rockets at the end of the war to make physical measurements up to unprecedented heights in the upper atmosphere. The programme, boldly conceived, ambitious in scope, consumptive of much time, effort and money, has suffered many setbacks and disappointments, but is now beginning to achieve some very interesting and new results unobtainable by other methods.

The first section of the book, which deals with rocket techniques, gives some idea of the great technical difficulties involved in obtaining reliable, representative readings from sensitive instruments when mounted on such an unstable platform and subjected to very large accelerations.

In the next section are described the various methods used in measuring temperature, pressure and wind in the upper atmosphere—occasionally up to heights of 220 km. Here some authors give the impression of being far more concerned with techniques than with the results. A careful assessment of the data obtained by the various methods and more discussion of the discrepancies would have been most valuable.

The discussion on the composition of the high atmosphere produced some interesting contributions. It appears that sampling from rockets reveals no diffusive separation of gases up to heights of 60 km. At higher levels, though the analysis would appear to point to some gravitational separation of argon from neon and helium, there is reason to believe that the result is spurious and arises from a defect in the sampling system. Measurements with a radio-frequency mass spectrometer fail to indicate separation up to 140 km.

Rocket-borne instruments have been used to explore the ionosphere and to measure the effective electron density, collision frequency and to investigate propagation modes over long paths. Interesting data on the electron density as a function of altitude in the E layer have been obtained by Lien *et al.*

Among other important contributions, one may mention the use by Johnson, Purcell *et al.* of a grating spectrograph to obtain highly resolved solar spectra down to wavelengths of 2100Å, and the measurement of the intensity and composition of the primary cosmic radiation by Johnson, Davis and Siry.

There can be little doubt that the rocket has proved an important new research tool in atmospheric physics. But so far, very few soundings have been made outside New Mexico and the danger of inferring too much about atmospheric behaviour from measurements made at one point must always be borne in mind. The enormous cost of obtaining data on a synoptic basis would appear prohibitive, but some encouragement is given in this book by van Allen and Gottlieb who propose that a small instrumental rocket be lifted into the lower stratosphere by a balloon and then fired in a nearly vertical direction, thus eliminating the aerodynamic resistance of the lower atmosphere. Five successful launchings of this type in which payloads of 30 pounds were sent to altitudes of 64 miles have been made at a cost of 2000 dollars each, which may be compared with nearly half a million dollars involved in the launching of a U.S. Viking rocket.

In the published proceedings of a conference one must expect an uneven standard of subject matter and presentation. Some of the papers in this volume are admirably concise and well written, others could hardly be worse. Had the editors insisted on the elimination of all repetition, overlap and unnecessary detail, they could have produced a considerably smaller volume without sacrifice of scientific content, but, of course, only at the expense of considerable delay. The present book is handsomely produced, but one wonders whether something cheaper would have been more appropriate, in view of the preliminary character of the work, the results of which may soon be superseded. B. J. MASON.

CONTENTS OF SECTION A

PAGE

Dr. G. ABRAHAM, Dr. L. COHEN and Dr. A. S. ROBERTS. The Binding Energies of ^{3}H and ^{4}He 265

Dr. W. J. SWIATECKI. The Effect of a Potential Gradient on the Density of a Degenerate Fermi Gas 285

Mr. R. C. SRIVASTAVA. Performance of Hot Wire Thermal Diffusion Columns 294

Mr. E. HOLÖIEN. Note on a Search for Simple Analytic Wave Functions for Configurations $(1s)^2(2s)$ and $(1s)^2(2s)^2$ in Atoms and Ions from He to C . 297

Dr. K. L. ERDMAN. The Angular Correlation of Annihilation Radiation . . 304

Mr. J. A. BASTIN and Dr. R. W. WRIGHT. The Electrical Conductivity of Cadmium Oxide at Low Temperatures 312

Dr. A. H. DE BORDE. The Effect of Radiative Corrections on a Charged Spin 1/2 Particle in a Constant Magnetic Field 316

Mr. D. J. CANDLIN. On the Identification of X-Ray Satellites . . . 322

Mr. J. N. MURRELL and Prof. H. C. LONGUET-HIGGINS. The Electronic Spectra of Aromatic Molecules: III—The Effect of Inductive Substituents . . 329

Research Notes:

Dr. C. B. O. MOHR. The Mean Free Path of Slow Neutrons in Nuclear Matter 340
Dr. C. B. O. MOHR. Positronium Formation in Gases and its Pressure Dependence 342
Prof. D. R. BATES, Charge Transfer and Ion–Atom Interchange Collisions . 344
Dr. E. W. EMERY and Mr. N. VEALL. The Branching Ratios of ^{86}Rb and ^{42}K 346

Letters to the Editor:

Dr. H. N. YADAV. The Nuclear Scattering of Electrons by Uranium 348
Mr. T. P. GRAY and Dr. G. R. SATCHLER. Angular Correlation Parameters for Oriented Nuclei 349
Miss M. C. M. O'BRIEN. The Magnetic Ground State of UF_3 . . 351
Dr. J. B. HARDING. The Second Maxima in the Transition Curve of Cosmic Ray Showers 352
Prof. H. T. FLINT and Miss E. M. WILLIAMSON. A Note on the Relativistic Cut-Off 354

Reviews of Books 356

Contents of Section B 360

Recovery and Recrystallization in Metals examined in terms of the Restoration of the Transient Creep Properties

By A. J. KENNEDY

The British Iron and Steel Research Association, London S.W.11

MS. received 1*st December* 1954 *and in amended form* 28*th December* 1954

Abstract. Pure lead wires extended by creep under identical conditions have been subjected to a period of anneal, during which the stress was removed, and then re-tested under the original creep conditions. The effect of the temperature, the time of the anneal, and the grain size on the subsequent creep properties have been examined. As far as the grain size is concerned, there appears to be very little effect until the grain diameter becomes commensurate with the specimen diameter, when the amount of recovery softening, under any particular conditions, increases. The general features of interrupted creep observed previously by the author in leads which did not recrystallize are again exhibited, in particular, the creep results can be fitted by assuming some fraction N of the material to have had its original creep properties fully restored. This fraction is now made up of both recovered, n, and recrystallized, X, material. The time function of the recovery process is the same as that observed in the non-recrystallizing leads, and it is thus possible to compute the values of n and X in any particular case. A theory of recrystallization is developed based on the concept that nuclei are formed by the recovery process; this is shown to satisfy the observations and provides an activation energy for the growth of new material of 31 000 cal mole^{-1}. The activation energy for recovery is found to be 19 800 cal mole^{-1} and that for creep 18 400 cal mole^{-1}. Some consequences of this approach are reviewed in the light of previous work.

§1. INTRODUCTION

The creep behaviour of metals subjected to interrupted stressing has been examined fairly extensively (Kennedy 1952, 1953 a). The main studies have been directed towards solving the problem of the creep behaviour of a metal subjected to pulses of constant stress under isothermal conditions. It has been shown that the type of phenomena observed can be interpreted on the basis that a certain fraction of the metal recovers in full its transient creep properties: the variation of this fraction with the load time and the off-load time has been established, and certain consequences of the mathematical treatment worked out and confirmed experimentally. Possibly the most notable fact is that it is possible to achieve a greater creep strain under interrupted conditions than under continuous stress for the same total test time. Further experiment (Kennedy 1953 b) has shown that if the stress is reduced during the ' off-load ' period, as distinct from being removed altogether, the amount of recovery is less, being greatest at zero stress. It is also clear that grain size has an influence on the amount of recovery that occurs and is particularly important when the grain dimensions approach that of the specimen cross section. The extreme case is that of a single crystal, which shows a very rapid recovery (Slifkin and Kauzmann 1952) compared with that of the polycrystalline material. It would appear valuable, then, to examine more closely the dependence of recovery on grain size.

In the above interrupted creep work, the restoration of the creep properties has been used as an index of the amount of recovery which has taken place. A great deal of work has already been done, of course, on the recovery of other properties during annealing : for instance, electrical conductivity (Manintveld 1952, Druyvesteyn and Manintveld 1951) hardness (Wilson and Thomassen 1934) and modulus (Smith 1953) with particular attention being paid to the temperature dependence of the processes involved. As a result of such work, the main features of recovery have been reconciled with dislocation theory (Burgers 1949, Mott 1951, Cottrell 1955), notably the phenomenon of polygonization (Cahn 1948, 1949).

As a new approach to the recovery and recrystallization problem, it appeared valuable to attempt an extension of the original interrupted creep work by determining the dependence of the recovery on the temperature of the specimen during the recovery period. Such work would not only yield fundamental information on the recovery mechanism itself, but would also constitute a first step towards the solution of the important applied problem of the creep response of metals in structures and machines when both stress and temperature are varied. By using a pure lead, which recrystallizes under suitable conditions, it has also been possible to demonstrate the difference in behaviour of this material compared with that of a non-recrystallizing lead. As recrystallization brings about a new mean grain size, the dependence of recovery on this factor has also been examined. On the basis of the observations, a new treatment of the nucleation problem is proposed: this clarifies the relationship between recovery and recrystallization and accords with much of the work previously reported on the subject.

§ 2. COMPARATIVE TESTS ON RECRYSTALLIZING AND NON-RECRYSTALLIZING LEADS AND THE EFFECT OF GRAIN SIZE

The leads used in these series of experiments may be conveniently designated 'recrystallizing' and 'non-recrystallizing', such terms being taken to apply only over the range of experimental conditions used. The 'non-recrystallizing' materials certainly recrystallize if strained at a sufficiently high temperature.

The recrystallizing leads used were (*a*) super-purity Tadanac lead (99·999%) supplied by British Non-Ferrous Metals Research Association, and (*b*) a less pure lead (99·99%) extruded for us by Associated Lead Manufacturers Ltd. The non-recrystallizing leads were (*a*) a lead containing 0·05% tellurium supplied by British Non-Ferrous Metals Research Association and (*b*) a commercial lead (99·0%). The non-recrystallizing leads are of the same composition as those previously used in the original interrupted creep work (Kennedy 1952) and a short summary of the analysis of this work is interposed here.

Suppose the transient creep curve of a metal under constant stress can be represented by an equation of the form

$$\epsilon = \epsilon_0(1+\beta t^p) \qquad \ldots\ldots(1)$$

where ϵ is the length of the specimen at a time t after the imposition of the stress, and ϵ_0, β and p are constants. If the stress is interrupted after a time τ, and the stress removed for a period of recovery R, then the curve on reloading is given by the sum of two components : (i) that exhibited by a fraction n of the material which behaves as if its original transient creep properties were fully restored, and

(ii) that exhibited by the remaining fraction $1-n$ which behaves as if unloading had not occurred. Thus the form of the subsequent curve is given by

$$\epsilon=\epsilon_0\{n\beta(t-\tau)^p+(1-n)\beta(t^p-\tau^p)+(1+\beta\tau^p)\}. \qquad \ldots\ldots(2)$$

The dependence of n on both τ and R has been investigated in commercial leads and in a lead containing 0·05% tellurium for which $p=\frac{1}{3}$ under the particular conditions of the test; thus equation (1) above becomes the transient creep equation of Andrade (1910). The recovery temperature in these original tests was the same as the creep temperature, and under these conditions the materials showed no recrystallization.

It was demonstrated, in particular, that the expression

$$(1-n)/n=A(\tau/R)^{1/3} \qquad \ldots\ldots(3)$$

where A is a constant, applied unless R exceeded some critical value R_c in which case the fraction n recovered was less than given by equation (3). The constant may be said to characterize the recovery properties of the material. High values of A, the recovery constant, imply a low recovery rate, and conversely. It is an empirical fact that the value of A is largely independent of the test conditions. It should be noted that this formulation of the interrupted creep behaviour of metals is the simplest that could be found to fit the observations and that no other treatment of the creep and recovery problem has been proposed. It can be readily demonstrated that recovery, when examined in this way, is not characterized by a logarithmic law, such as that proposed as the relation between x-ray line breadth and time (Paterson and Orowan 1948). The work in this paper is based upon the empirical validity of the foregoing analysis as far as non-recrystallizing materials are concerned.

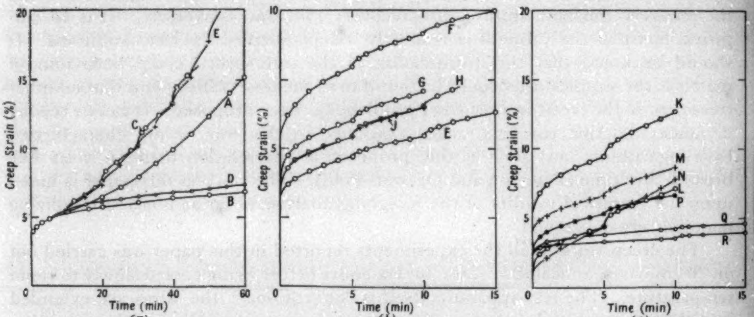
The creep work in all the experiments reported in this paper was carried out on 20 cm wires, annealed at 120°c for six hours before being cooled slowly to room temperature. The test apparatus itself is conventional: the wires are extended in a thermostatic enclosure by means of a constant stress device, the extension being measured by a travelling microscope reading to 0·005 mm.

The influence of recrystallization on the general form of the creep curve is well known (Sully, Cale and Willoughby 1948, Andrade 1948). The results shown in figure 1 compare the behaviour of the Tadanac and the commercial leads (curves A and B) under conditions such that each exhibits the same initial transient properties (that is, the early stages of the two curves coincide). The general character of the difference is what one would expect.

We now subject each to cycles of interrupted stressing in which the stress is applied for 10 minutes and removed for 10 minutes. In the case of the tellurium lead (curve D), the result is in accord with the treatment already proposed. The super-purity lead (curve C), however, shows a markedly higher 'recovery', taking as a measure the magnitude of the new transient exhibited on reloading. A further test on this material (curve E) under cycles of five minutes load and five minutes off load (i.e. the same on–off ratio) is also shown. There is no distinctive difference between the two results as far as the strain achieved is concerned. As has been shown for the non-recrystallizing material, the final strain depends on the stress on–off ratio rather than on the frequency. We note also that there is a significant instantaneous extension of the Tadanac lead on reloading—an effect not so far observed at all with the non-recrystallizing materials, although it is possible that such a phenomenon might be observed under extreme conditions.

The question which at once arises is how much of this apparent 'recovery', or restoration, of transient creep is due to true recovery (involving no structural change) and how much is due to recrystallization (involving the nucleation and growth of new stress-free grains). It is clear that if true recovery is a function of the grain size (which is changing throughout the test), and if recrystallization itself is in any way dependent upon the progress of the recovery process, then the whole problem is a very complex one.

We first attempt to examine how the grain size affects the rate of recovery. The Tadanac lead recrystallizes under the creep conditions of figure 1 (*a*), so that if such a test is stopped at some point, and the metal annealed, a new mean grain size will be obtained. In figure 1 (*b*) the results are shown of creep tests on the different materials obtained by such pre-straining treatments. Curve E is the creep curve for the original material, with a mean grain diameter of 0·12 mm (the wire diameter is 2 mm), and curves J, H, G and F are for materials subjected to progressively greater strain, under creep conditions, prior to a final anneal at 110°c for 24 hours. The grain sizes lie in the ranges indicated in the table, which also sets out the results of analysing the test results. It will be seen that although the linear grain size increased by a factor of about 10, the fraction recovered *n*, and hence the recovery constant *A*, showed no significant trends. A large number of tests on polycrystalline lead (including those previously published (Kennedy 1952)) indicate an average value for *A* of about 9·0, and this is very near to the *A*-values here observed. It will be noted that the β-values increase progressively with the creep strain imposed prior to the anneal, and may be taken to indicate a progressive increase in grain size. The fact that the ratio of the grain to the specimen diameter can affect the transient creep in this way has already been demonstrated (Andrade and Kennedy 1951). The tests in figure 1 (*b*), then, show that a relatively large change in the grain size has little effect on the recovery properties. Yet it is highly likely, if Slifkin and Kauzmann's experiments (1952) are at all representative, that single crystals exhibit a very much higher rate of recovery than the polycrystalline material. It appeared desirable, therefore, to extend the tests of figure 1 (*b*) by working with materials containing very large



(*a*) (*b*) (*c*)

Figure 1. Interrupted creep curves of recrystallizing and non-recrystallizing leads. Details of the curves are given in the table.

Curve	Lead	Test conditions				β (min$^{-1/3}$)	ϵ (cm)	n	Grain diam d (mm)	A	Notes
		Temp (°c)	Stress (kg cm^{-2})	τ (min)	R (min)						
A	Tad.	36	71	Continuous		$2{\cdot}08\times10^{-2}$	21·1	—	—	—	
B	Te	36	80	Continuous		$2{\cdot}08\times10^{-2}$	21·1	—	—	—	
C	Tad.	36	71	10	10	$2{\cdot}08\times10^{-2}$	21·1	0·11	0·12	8·0	No pre-strain. Annealed 110°c for 1 h.
D	Te	36	80	10	10	$2{\cdot}08\times10^{-2}$	21·1	—	—	—	
E	Tad.	36	71	5	5	$2{\cdot}08\times10^{-2}$	21·1	0·11	0·12	8·0	Pre-strained by creep and annealed prior to test.
F	Tad.	36	71	5	5	$3{\cdot}62\times10^{-2}$	21·8	0·078	1·0–1·2	11·7	
G	Tad.	36	71	5	5	$2{\cdot}38\times10^{-2}$	21·5	0·11	0·6–0·8	8·3	
H	Tad.	36	71	5	5	$2{\cdot}16\times10^{-2}$	21·4	0·12	0·6–0·8	7·1	
J	Tad.	36	71	5	5	$2{\cdot}20\times10^{-2}$	21·4	0·12	0·6–0·8	7·7	
K	Tad.	60	53	5	6	$2{\cdot}54\times10^{-2}$	10·5	0·38	4–10+	1·7	Successively strained and heated to produce large-grained specimens.
L	Tad.	60	53	5	6	$2{\cdot}03\times10^{-2}$	10·3	0·45	2–4	1·3	
M	Tad.	25	63	5	5	$1{\cdot}56\times10^{-2}$	10·2	0·45	2–3	3·1	
N	Tad.	25	76	5	5	$2{\cdot}08\times10^{-2}$	10·1	0·25	1–3	3·0	
P	Tad.	25	63	5	5	$3{\cdot}45\times10^{-2}$	9·9	0·22	2–3	3·5	
Q	Te	36	60	10	45	$7{\cdot}75\times10^{-3}$	10·3	0·24	0·20	6·7	
R	Te	36	60	10	45	$4{\cdot}40\times10^{-3}$	10·3	0·63	0·40	4·5	

grains to see if a trend towards this higher recovery rate was exhibited. These large-grained materials were prepared by successively straining and heating in the standard way.

In figure 1 (*c*) some of the results of these tests are plotted, showing the behaviour of the materials under different conditions, the curves being analysed in the table. Again, only a range of mean values of the grain diameters is given, as the grain size was extremely non-uniform. In the case of K, for instance, some grains extended for about 2 cm along the wire. The broad result of these tests is to show that much higher rates of recovery (that is, lower A-values) can be achieved if the grain size is made large enough compared with the specimen diameter. In general, the results of the table indicate that grain size does not affect the recovery properties of the metal until the grain diameter d becomes commensurate with that of the specimen, when a relatively sudden increase in the rate of recovery occurs. No more exact description of the (A, d) relation is possible until tests can be made on very uniform aggregates consisting of a few grains only.

§ 3. RECOVERY TESTS ON RECRYSTALLIZING LEAD

A series of experiments was next conducted on the pure lead supplied to us by Associated Lead Manufacturers Ltd. The object of these fairly extensive tests was to establish the nature of the difference in the behaviour of recrystallizing and non-recrystallizing metals, when each is subjected to interrupted creep. The behaviour of the non-recrystallizing lead has already been examined and analysed (Kennedy 1952).

The present work was carried out on 20 cm wires 1·5 mm in diameter, annealed at 120°c for 24 hours before being cooled slowly to room temperature for testing under constant stress. The first part of the test, for all the wires examined, consisted of an extension under a constant stress of 112 kg cm^{-2} at 22°c for five minutes. The stress was then removed and the wire under test placed in a constant temperature enclosure set at the desired ‘ recovery temperature ’ for a chosen period. At the end of this period it was replaced in the extension apparatus, where its temperature returned to 22°c, and then extended under the same constant stress. The creep extensions were observed by a travelling microscope reading to 0·005 mm. The results obtained from a series of such experiments are shown in figures 2 and 3. Each set of results shows the effect of recovery at different temperatures for the same time. In addition to those reproduced a few tests were carried out for very short recovery times. There is clearly a limit to the shortness of this period if the temperature changes involved take a finite time to complete—as they do when the specimen is heated or cooled in an air enclosure. The results will also be less accurate, as the cooling and heating periods take up a significant part of the total recovery time. The results for these cases were obtained, in fact, by using liquid baths, and by restricting the temperature change to a fairly small range (about 30°c).

§ 4. INTERPRETATION OF RESULTS

The first step in the examination of these results was to fit the observations according to the treatment already proposed (Kennedy 1952).

The curves plotted in figures 3 and 4 were fitted to expression (2), and the values of n deduced. It was found that, at the higher temperatures, in addition

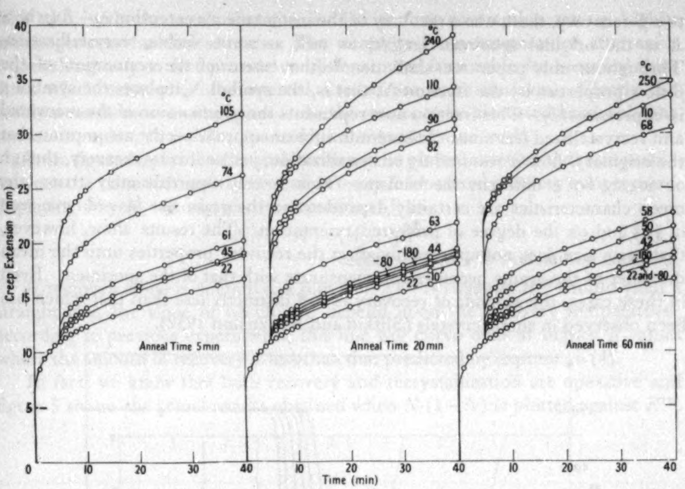


Figure 2. Creep curves of a pure lead under a stress of 112 kg cm^{-2} at 22°c, interrupted after 5 minutes creep and annealed for the times shown. The temperature of the anneal is given against the subsequent creep curve.

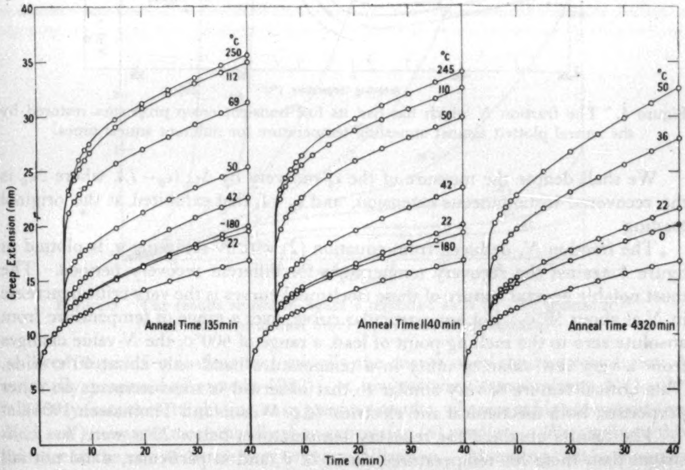


Figure 3. Creep curves of a pure lead under a stress of 112 kg cm^{-2} at 22°c, interrupted after 5 minutes creep and annealed for the times shown. The temperature of the anneal is given against the subsequent creep curve.

to a β-recovery, there was a recovery of the instantaneous extension $\epsilon_0 - L$ (where L is the original specimen length) as well as some visible recrystallization. Throughout this paper we shall denote the extent of the restoration of the β-flow properties by the fraction N, that is, the symbol N replaces the symbol n in expression (2). This fraction now represents the *effective* sum of the recovered and recrystallized fractions. Inherent in such an approach is the assumption that the original β-flow is restored by recrystallization, just as it is by recovery, though obviously by a different mechanism. This is only approximately true: the creep characteristics are certainly dependent on the grain size (as we have seen in § 2) and on the degree of preferred orientation. The results show, however, that grain size does not appreciably affect the recovery properties until the mean diameter of the grains becomes commensurate with that of the specimen. Even in these cases, the amount of recovery is still distinctly less than that which has been observed in single crystals (Slifkin and Kauzmann 1952).

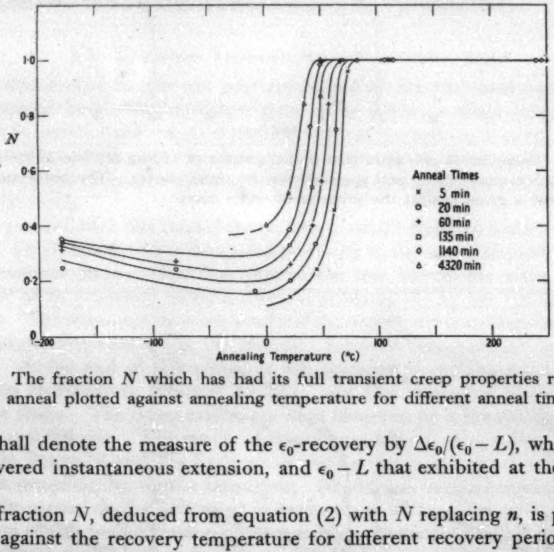


Figure 4. The fraction N which has had its full transient creep properties restored by the anneal plotted against annealing temperature for different anneal times.

We shall denote the measure of the ϵ_0-recovery by $\Delta\epsilon_0/(\epsilon_0 - L)$, where $\Delta\epsilon_0$ is the recovered instantaneous extension, and $\epsilon_0 - L$ that exhibited at the original loading.

The fraction N, deduced from equation (2) with N replacing n, is plotted in figure 4 against the recovery temperature for different recovery periods. The most notable general feature of these isochronal curves is the very critical increase in N at about 50°c. For any particular curve, over a range of temperature from absolute zero to the melting point of lead, a range of 600°c, the N-value changes from a very low value to unity in a temperature band only about 60°c wide. This critical feature is very similar to that observed in measurements on other properties, both mechanical and electrical (e.g. Wilson and Thomassen 1934).

The results obtained for recovery temperatures below 22°c were less consistent than those for temperatures above 22°c, and in particular, n did not fall towards zero, but remained roughly constant at about 0·3. Indeed, there appeared to be a tendency for n to increase as the temperature was reduced. This is possibly due to the fact that cooling was effected very rapidly by plunging the wire into a cold bath, with the result that the sudden surface contraction set up highly inhomogeneous internal stresses. The return to room temperature after such a quenching may in fact provide conditions more favourable for recovery than those which exist when the internal stresses are more uniform. Clearly the problem needs more examination. In the results to be discussed here, only the consistent region, above 0°c, was used in the quantitative work.

From figure 4 a set of isothermal curves can be plotted showing the relationship between the value of N and the time of recovery R for a fixed recovery temperature. Replacing n by N, we may rewrite expression (3) as

$$N/(1-N) = (A/\tau^{1/3})R^{1/3} \qquad \ldots\ldots(4)$$
$$= \text{const.} \times R^{1/3} \text{ if } \tau \text{ is maintained constant.}$$

If recovery alone is operative, a plot of $N/(1-N)$ against $R^{1/3}$ should yield a straight line, the slope of which will depend upon the recovery temperature; according to previous experiments, this line will curve over at higher R-values where the amount of recovery is less than that predicted by expression (4).

In fact, we know that both recovery and recrystallization are operative and figure 5 shows the actual results obtained when $N/(1-N)$ is plotted against $R^{1/3}$.

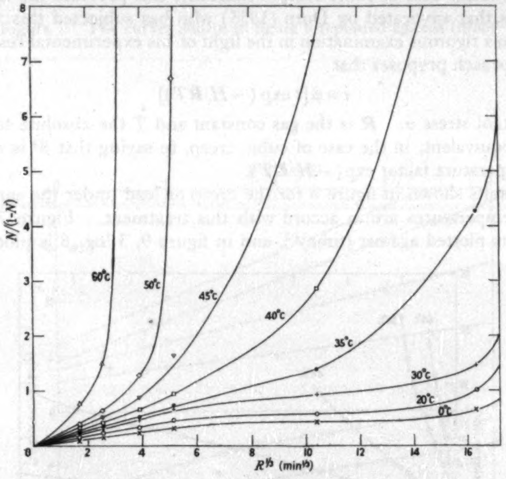


Figure 5. The results shown in figure 4 replotted as $N/(1-N)$ against $R^{1/3}$ (where R is the anneal time) for various temperatures.

From figure 5 it is clear that initially the curves obey the recovery function originally proposed (Kennedy 1952); at low recovery temperatures $N/(1-N)$ falls away from the straight line, as in previous results, whereas at high temperatures $N/(1-N)$ increases more rapidly than expression (4) predicts. We attribute this increase in the rate of N-recovery to recrystallization. Assuming that in the early stages the fraction is wholly due to recovery, then in this range we may therefore substitute n, the recovered fraction, for N in equation (4). The results cannot be developed any further without making certain assumptions. The best that can be done is to maintain the established form of the recrystallization problem as being a process of nucleation and growth, and to select the most reasonable assumptions on the evidence at present available. The predictions, or results, of such a treatment may then be compared with other work. The important point in the following, it is suggested, is not the particular quantitative values that are obtained, but rather the principle that recovery can be introduced into recrystallization theory as a nucleus-creating process, and that such a treatment is in harmony with much of the experimental evidence, in particular with that reported in this paper.

§ 5. Temperature Dependence of Creep and Recovery

The first question which arises in dealing with these results is how temperature is to be introduced into the recovery process represented by expression (4). Noting the parallel which exists between (4) and the creep equation $\epsilon=\beta t^{1/3}$ one considers how temperature is introduced into transient creep theories.

Although several forms of temperature dependence have been proposed for transient creep, the most acceptable appears to be that adopted by Mott (1953) for the particular case of cubic creep. Basically, this provides the same type of function as that advocated by Dorn (1955) who has subjected this type of formulation to a rigorous examination in the light of his experimental results. This type of approach proposes that

$$\epsilon=\psi[t\exp(-H/\boldsymbol{R}T)] \qquad \ldots\ldots(5)$$

for a constant stress σ. $\boldsymbol{R}$ is the gas constant and T the absolute temperature. It is thus equivalent, in the case of cubic creep, to saying that β^3 is operated on by the temperature factor $\exp(-H/\boldsymbol{R}T)$.

The results shown in figure 6 for the creep of lead under the same stress at different temperatures are in accord with this treatment. Figure 7 shows the same results plotted against (time)$^{1/3}$, and in figure 9, 3 $\log_{10}\beta$ is plotted against

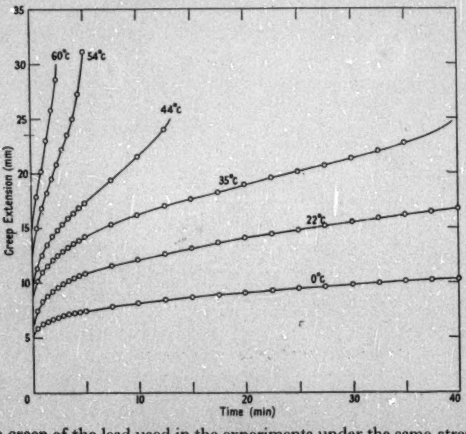


Figure 6. The creep of the lead used in the experiments under the same stress 112 kg cm^{-2} at different temperatures.

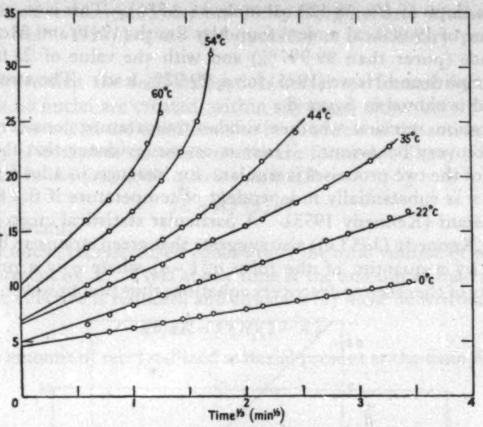


Figure 7. The curves shown in figure 6 replotted against $(\text{time})^{1/3}$.

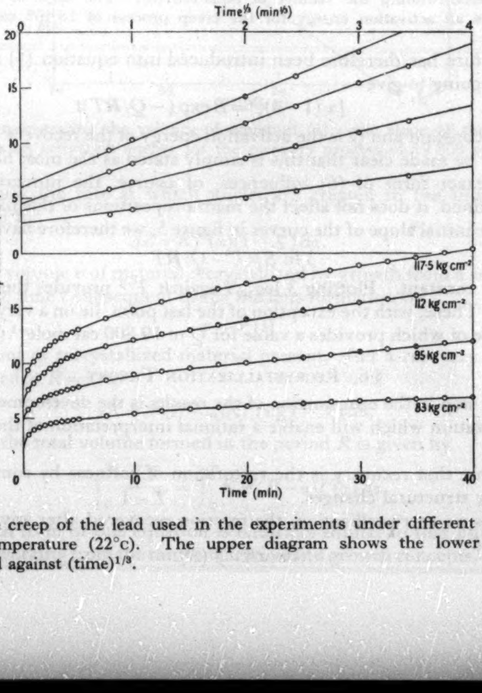


Figure 8. The creep of the lead used in the experiments under different stresses at the same temperature (22°c). The upper diagram shows the lower creep curves replotted against $(\text{time})^{1/3}$.

T^{-1} to give a slope of $H = 18\,400$ cal mole^{-1} ($\pm 5\%$). This is in good agreement with the value of 19 000 cal mole^{-1} found by Smith (1941) and McKeown (1937) for pure leads (purer than 99·997%) and with the value of 23 000 cal mole^{-1} obtained by Smith and Howe (1945) for a 99·92% lead. The stress response of the lead used is shown in figure 8.

The question remains whether similar temperature considerations can be applied to recovery behaviour. There is certain evidence that the temperature dependence of the two processes is similar: for instance, in a load–unload–reload experiment, n is substantially independent of temperature if the times involved are kept constant (Kennedy 1955). A particular statistical creep model (briefly indicated by Kennedy (1953 c)) also suggests that creep strain can be significantly represented by a quantity of the form $q/(1-q)$ where q is a proper fraction. The similarity of the creep and recovery equations thus becomes even more striking.

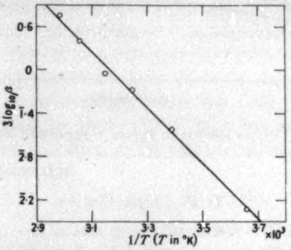


Figure 9. Demonstrating the validity of equation (5). The slope of the line drawn provides an activation energy for the creep process of 18 400 cal mole^{-1}.

Temperature has therefore been introduced into equation (4) in accordance with the foregoing to give

$$[n/(1-n)]^3 = B \exp(-Q/\boldsymbol{R}T)t \qquad \ldots\ldots(6)$$

where B is a constant and Q is the activation energy of the recovery process.

It should be made clear that this is simply stated as the most likely equation. Whilst the exact form of (6) influences, of course, the numerical activation energies obtained, it does not affect the main propositions of this paper.

If S is the initial slope of the curves in figure 5, we therefore have

$$3 \ln S = C - Q/\boldsymbol{R}T \qquad \ldots\ldots(7)$$

where C is a constant. Plotting $3 \log_{10} S$ against T^{-1} provides the points shown in figure 10. These, with the exception of the last point, lie on a very good straight line, the slope of which provides a value for Q of 19 800 cal mole^{-1} ($\pm 5\%$).

§ 6. RECRYSTALLIZATION THEORY

The next step in the examination of the results is the development of a theory of recrystallization which will enable a rational interpretation of the results to be made.

We assume that recovery is the restoration of softness by some mechanism not involving structural changes.

If there is no recrystallization, the fraction recovered after any time is given by $n = f(R)$ or

$$dn/dR = f'(R). \qquad \ldots\ldots(8)$$

From empirical evidence (Yokobori 1951, Burghoff and Mathewson 1941, Gensamer and Mehl 1938) we say that recovery is a necessary prelude to recrystallization.

Let us suppose that the recrystallization nuclei are formed *by* the recovery process. Thus no nuclei are created, within any zone, either before or after that zone recovers; the nuclei are considered to be products of the action of recovery. When recovery is complete no more nuclei are formed. Then, if the number of nuclei formed in the time from α to $\alpha + d\alpha$ is dZ

$$dZ = Kf'(\alpha)\, d\alpha \qquad \text{......(9)}$$

where K is a constant.

From these nuclei recrystallized zones grow, the total volume of which, at any instant, we represent by X. Thus at any time R the amount of material in which recovery can be effective is reduced, and equation (8) must be written as

$$dn/dR = f'(R)(1 - X) \qquad \text{......(10)}$$

where X is the amount of recrystallized material present at the time R.

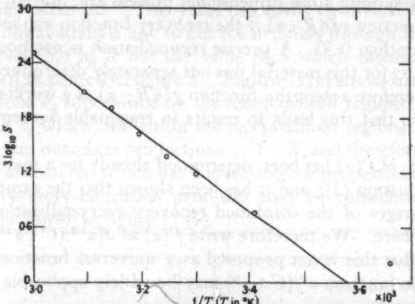


Figure 10. Demonstrating the validity of equation (7): the slope of the straight line drawn gives an activation energy for the recovery process of 19 800 cal mole^{-1}.

Consequently the rate at which nuclei are created in the unrecrystallized matrix is given by

$$dZ = Kf'(\alpha)(1 - X)\, d\alpha. \qquad \text{......(11)}$$

Now if the volume v of material recrystallized by growth from a single nucleus is related to the time t subsequent to the nucleus formation by

$$v = \mathscr{F}(t) \qquad \text{......(12)}$$

then the amount of recrystallized material present, after a recovery time R, due to nuclei formed at $R = \alpha$, is

$$dX = \mathscr{F}(R - \alpha)Kf'(\alpha)(1 - X)\, d\alpha. \qquad \text{......(13)}$$

Therefore the total volume formed in the period R is given by

$$\int_0^X \frac{dX}{1 - X} = \int_0^R K\mathscr{F}(R - \alpha)f'(\alpha)\, d\alpha. \qquad \text{......(14)}$$

The general form of this equation is naturally similar to that put forward by other authors working from the same nucleation and growth concepts, but recovery is here specifically introduced and replaces the arbitrary nucleation function of Johnson and Mehl (1939) and the exponential nucleation function of Avrami (1939, 1940, 1941). It differs fundamentally from Cook and Richards' approach (1947) in that the growth of new grains is not restricted to the recovered regions only. Three major considerations remain. These are: (*a*) the form of $\mathscr{F}(R-\alpha)$, (*b*) the form of $f'(\alpha)$ and (*c*) the limitation to growth imposed by the impingement of grains. We shall discuss these points separately.

(*a*) In a medium in which growth is uniform in all three dimensions $\mathscr{F}(R-\alpha)$ should be of the form $gD(R-\alpha)^3$ where D is a shape factor and g the linear velocity of growth: in two dimensions this becomes $gD(R-\alpha)^2$ and in one dimension $gD(R-\alpha)$. It is an experimental fact that growth is not uniform in three dimensions but is a function of the relative orientation of adjacent crystals. Further, when the specimen is in the form of a wire, as it is in these experiments, crystal growth in directions normal to the specimen axis is restricted, and ultimately growth proceeds almost wholly in an axial direction as a one-dimensional process. Measurements made by Anderson and Mehl (1945) on the fraction recrystallized in aluminium assuming a *two*-dimensional process are, indeed, not incompatible with a linear function $gD(R-\alpha)$ if the recovery function exponent is taken to be about 0·6 rather than 0·33. A precise reconciliation is not possible because the recovery function for this material was not separately determined.

We shall therefore accept the function $gD(R-\alpha)$ as a working hypothesis and it will be shown that this leads to results in reasonable agreement with the observations.

(*b*) The form of $f'(\alpha)$ has been determined already for a material that does not recrystallize (equation (3)), and it has been shown that the same function applies to the earlier stages of the combined recovery–recrystallization process in the pure lead used here. We therefore write $f'(\alpha)$ as $d[\alpha^{1/3}/(C+\alpha^{1/3})]/d\alpha$. It should be made clear that this is not proposed as a universal function; it is, however, possible that the function $\alpha^p/(C+\alpha^p)$ may be widely applicable with p being the creep law exponent.

(*c*) The question of grain impingement is generally recognized as being extremely difficult to formulate with any exactness. In this work we shall deal only with that region where X is small enough (say less than 0·5) for impingement considerations to be ignored.

§ 7. APPLICATION OF THE THEORY TO THE EXPERIMENTAL RESULTS

In order to apply the theory to the experimental results, we convert equation (14) into a more practical form.

Integrating, we have

$$\ln[1/(1-X)] = gDK\left\{\Big[(R-\alpha)f(\alpha)\Big]_0^R + \int_0^R f(\alpha)\,d\alpha\right\}$$

$$= gDK\int_0^R f(\alpha)\,d\alpha.$$

If our simple link between recovery and nucleation is valid, K is a constant for fixed pre-straining conditions (as in these tests), being the constant of proportionality in equation (9). For instance, if the recovery of each element or zone of the matrix produced one nucleus, K would be unity. D, the shape factor, is also a constant.

Thus the temperature dependence of gDK is simply that of g, the linear growth rate of the recrystallized material. Embodying the constants, we can therefore write gDK as $G \exp(-P/\boldsymbol{R}T)$.

We thus have

$$\ln [1/(1-X)] = G \exp(-P/\boldsymbol{R}T) \int_0^R [\alpha^{1/3}/(C+\alpha^{1/3})]\, d\alpha \quad \ldots\ldots(15)$$

where G is a constant and P the activation energy of the recrystallization process. The expression within the integral is, of course, temperature dependent and must be treated according to the previous recovery propositions.

Thus

$$\ln [1/(1-X)] = 3G \exp(-P/\boldsymbol{R}T)[\tfrac{1}{3}(R^{1/3}+C)^3 - \tfrac{3}{2}C(R^{1/3}+C)^2 + 3C^2(R^{1/3}+C) - C^3 \ln(R^{1/3}+C) - C^3(1{\cdot}83 - \ln C)]. \quad \ldots .(16)$$

In order to compute the right-hand side of this equation, it is convenient to use the fact that $n_0/(1-n_0) = R^{1/3}/C$ where n_0 is the value of the recovered fraction n according to equation (3), after a time R. That is, it is the value of n if no recrystallization had occurred. This value can easily be deduced from figure 5 by extending the initial straight line to cut the ordinate through $R^{1/3}$. It should be noted, of course, that n_0 is not the value of n which satisfies the quantity $(X+n)/[1-(X+n)]$ as plotted in figure 5. Figure 11 (*a*) shows diagrammatically the recovered regions n_0 superposed on the recrystallized regions X. Assuming that the amount of n_0 which lies within the recrystallized regions is proportional to X, then the amount outside is proportional to $1-X$, and therefore $n = n_0(1-X)$. Thus the amount of material which has had its original properties restored by the combined recovery–recrystallization processes may be considered as made up of the sum $X+n$ as shown in figure 11 (*b*).

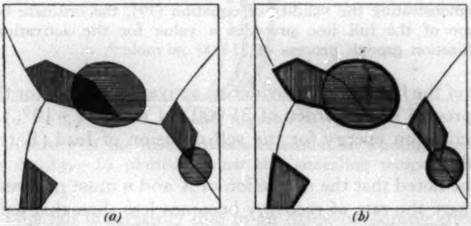


Figure 11. (*a*) The overlapping of the recovered regions n_0 (vertical hatching), with the recrystallized regions X (horizontal hatching). We may consider this to be made up, as shown in (*b*), by $X+n$, where $n = n_0(1-X)$. The vertical hatching now represents the quantity n. The region within the thick line N has thus had its full creep properties restored, whilst that outside remains unchanged.

By using the quantity n_0, and neglecting the small final term, equation (16) may be reduced to

$$\ln \{1/(1-X)\} = 3G \exp(-P/\boldsymbol{R}T) \times \left[\frac{R}{n_0^3}\left\{\frac{1}{3} - \frac{3}{2}(1-n_0) + 3(1-n_0)^2 - (1-n_0)^3 \ln\left(\frac{R^{1/3}}{n_0}\right)\right\}\right] \quad \ldots\ldots(17)$$

and the bracketed function on the right-hand side computed.

From the graphs plotted in figure 5 we can obtain the value of $N(=X+n)$ at any time R. As $N=X+n_0\,(1-X)$ we have

$$X=(N-n_0)(1-n_0) \qquad \ldots\ldots(18)$$

where both N and n_0 are known.

Thus, as equation (17) is equivalent to

$$\ln\left(\frac{1}{1-X}\right)\Big/F(R,n_0)=G\exp\left(-P/\boldsymbol{R}T\right)$$

we have

$$\ln\left[\ln\left(\frac{1}{1-X}\right)\Big/F(R,n_0)\right]=M-P/\boldsymbol{R}T \qquad \ldots\ldots(19)$$

where M is a constant.

By plotting the left-hand side against $1/T$, for various times R, and temperatures T, we obtain the results shown in figure 12.

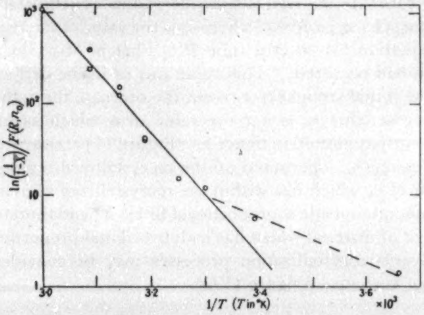


Figure 12. Demonstrating the validity of equation (19), the ordinate being logarithmic. The slope of the full line provides a value for the activation energy of the recrystallization growth process of 31 000 cal mole^{-1}.

The slope of the line drawn provides an activation energy for the recrystallization growth process of the order of 31 000 cal mole^{-1} ($\pm 10\%$). This is very near to the activation energy for the self-diffusion of lead (Barrer (1941) gives 27 900 cal mole^{-1}).

It should be noted that the derivation of X and n must proceed differently for those cases where the rate of recovery becomes less than that exhibited initially (the curves for 30°C, 20°C and 0°C in figure 5). The method assumes that the straight line region holds for a recovery period R_0: after this time it is assumed that no further recovery occurs (in fact recovery does occur but, as has been shown before, at a lower rate than that predicted by the initial recovery function). The increase in the value of $(X+n)/[1-(X+n)]$ after time R_0 is taken to be due solely to recrystallization by growth from the nuclei created up to the time R_0. Using the same thermal activation factor as appeared in equation (16), we have that the amount recrystallized, X, is related to the time R subsequent to R_0 by

$$X=GR\exp\left(-P/\boldsymbol{R}T\right)$$

where G is a constant, or

$$\ln(X/R)=M-P/\boldsymbol{R}T, \qquad \ldots\ldots(20)$$

M being a constant.

The expression ln (X/R) therefore replaces, for these cases, the left-hand side of equation (19).

The points deduced by both methods lie reasonably close to a common straight line (figure 12) although there is a deviation at higher values of $1/T$. Considering the approximations made in interpreting the curves obtained at the lower temperatures, and considering also that the smaller amount of recrystallization makes the fitting of the results less sensitive, it is thought that nothing significant can be deduced from the form of this deviation from linearity in figure 12. It is certainly possible, however, that a genuine change of slope does occur at lower temperatures.

§ 8. THE RECOVERY OF THE INSTANTANEOUS EXTENSION

The recovery of instantaneous extension, represented by using the quantity $\Delta\epsilon_0/(\epsilon_0 - L)$, is shown in figure 13 for different recovery temperatures. $\Delta\epsilon_0$ is the recovered instantaneous extension, and $\epsilon_0 - L$ the instantaneous extension at the original loading, L being the original specimen length. Under prolonged annealing at the higher temperatures, the value of $\Delta\epsilon_0$ rises to a maximum and then steadily falls. This fall is certainly due to an oxide-hardening effect, which can be large enough to prevent creep altogether under the original test conditions. Light etching of the surface restores in full the creep properties of the material. The curves of figure 13 are not regular enough to be examined in the way that the recovery–recrystallization curves of figure 4 have been, and consequently no accurate activation energy for the $\Delta\epsilon_0$-recovery can be deduced. It is evident, however, that the critical band of temperature, for the same range of recovery times, is somewhat higher than that of figure 4; alternatively, we may say that the $\Delta\epsilon_0$-recovery appears to be associated with the later stages of the recovery process, and not with the initial stages. It is suggested, therefore, that $\Delta\epsilon_0$-recovery is a result of recrystallization rather than of recovery, and that it is due to the instantaneous extension of the new strain-free crystals, inhibited, as they must be, by adjacent crystals of the original matrix.

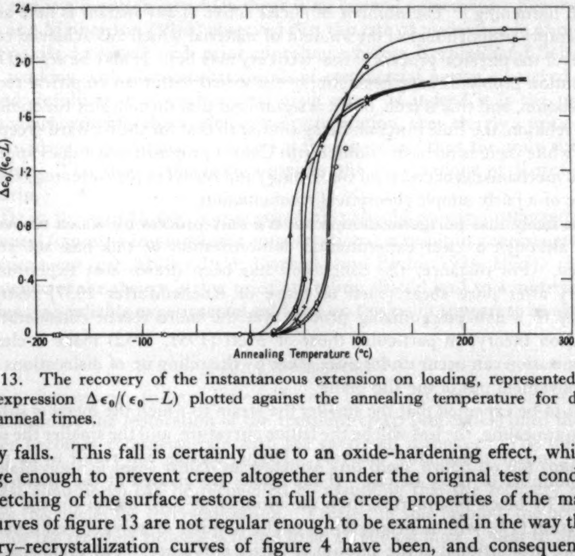


Figure 13. The recovery of the instantaneous extension on loading, represented by the expression $\Delta\epsilon_0/(\epsilon_0 - L)$ plotted against the annealing temperature for different anneal times.

§ 9. General Discussion and Summary of Results

The foregoing experimental results on the recovery–recrystallization processes are consistent with the view that the recovery process provides the recrystallization nuclei. The recovery process itself has been shown to follow the same time function as that obtained previously for a tellurium lead which did not recrystallize.

This view of the origin of the recrystallization nuclei is very similar, in principle, to that put forward by Cahn (1950). Cahn considers the nuclei to arise in the highly distorted regions (' local curvatures ') when these polygonize under annealing. The features of such a recrystallization model of this type, and their reconciliation with the published empirical data, have been so comprehensively reviewed by Cahn that it is pointless to retrace the arguments here in any detail.

It is important, however, to note in what respects the present work differs from that of Cahn. Whilst the present approach considers nuclei to be a product of a recovery process, and indicates that these nuclei may be logically expected to occur at a boundary between recovered and unrecovered material, it does not involve any assumptions regarding distributions of incubation times, or of degrees of local hardening. The number of nuclei active at any instant is here assumed to be simply proportional to the amount of material which has achieved recovery (whatever the physical process of that recovery may be). It may be argued that the distribution problems have been simply embodied within an empirical recovery–time relation, and this is true, but it is significant that no complex functions occur in this relation, the time function being similar to that for the forward creep curve. Thus, while there is no basic conflict with Cahn's propositions, it does appear as if the two mechanisms of creep (or hardening) and recovery (or softening) should be capable of a fairly simple theoretical reconciliation.

It is likely that polygonization is not the only process by which recovery can occur, although a clear experimental demonstration of this has not yet been achieved. For instance, the conclusion has been drawn that experiments on recovery after pure shear (such as those of Kochendörfer 1937) restrict the recovery to a non-polygonizing process, yet the more recent considerations of dislocation theory, in particular those of Mott (1951, 1952) make it clear how polygonization can occur under pure shear by the piling up of dislocations and by their movement out of the slip plane.

It is to be expected that the smaller the strain to which the metal is subjected, prior to annealing, the less will be the lattice curvature, and the smaller the number of polygonized blocks after recovery. Higher strains will result in an increase in the number of these blocks. If the nuclei are created at the discontinuities afforded by block boundaries, then the number of such nuclei (or the constant in equation (17)) will be proportional to the number of polygonized blocks, and therefore proportional to the amount of lattice curvature. The grain size, which is inversely proportional to the number of nuclei, will therefore increase as the pre-strain is diminished, and will be greatest for the minimum pre-strain necessary for recovery and recrystallization to occur. This agrees with the general experimental evidence (e.g. Eastwood, Bousu and Eddy 1935). A neat demonstration of the effect has been obtained by annealing tin sheets subjected to bullet impact: the results are reproduced by Doan and Mahla (1941).

The question of the dependence of the final grain size on the annealing temperature is less readily resolved. Previous work on α-brass (Eastwood, Bousu and Eddy 1935), aluminium (Anderson and Mehl 1945) and rock-salt (Müller 1935), has shown that for these materials the activation energy for nucleation (corresponding to the quantity Q in this paper) is very nearly the same as the activation energy for growth P, so that temperature had little effect on the final grain size. Certain deviations from this have been observed, notably at small prior strains, and in these cases Q was somewhat greater than P, so that the grain size increased with the annealing temperature.

If the problem of final grain size can be treated as simply as this, and the process of true grain growth (that is the growth of certain of the new grains at the expense of neighbours) ignored, then the results for lead presented in this paper suggest that the final grain size achieved by primary recrystallization will diminish with increasing temperature because Q (recovery) is less than P; actually $Q \sim 2P/3$. There is not yet sufficient evidence to verify this prediction. In any case, it is possible that P falls off at higher temperatures, as it does in copper (Karnop and Sachs 1930) and aluminium (comparing Kornfeld and Pavlov 1934, 1937 with Anderson and Mehl 1945).

The effect of subjecting the metal to recovery anneals prior to the recrystallization test is a question on which there is no general agreement. For instance, Collins and Mathewson (1940) observed that the rate of nucleation in aluminium single crystals decreased with prior annealing whereas Kornfeld and Schamarin (1937), working with aluminium crystals of commercial purity, observed that it increased, resulting in a smaller final grain size. If indeed it is the recovery process which provides the nuclei for recrystallization, then clearly a prior anneal must provide a larger number of nuclei at the outset and thus the grain size must be smaller. The theory therefore supports the observations of Kornfeld and Schamarin.

As far as the growth rate is concerned, this should remain unaffected by the prior anneal (ignoring impingement considerations), a result in general agreement with experiment (e.g. Müller 1935, Kornfeld and Pavlov 1934, 1937). Another consequence of the theory is that prior annealing should lead to a more uniform grain size; no reliable experimental evidence on this point appears to be available.

§ 10. CONCLUSIONS

Grain size appears to have little influence on the rate of recovery softening (as evidenced by the restoration of the transient creep properties) until the grain diameter becomes commensurate with the specimen diameter. The difference in the behaviour of leads which recrystallize and those which do not (under the same test conditions) can be reconciled by assuming that the recovery process provides the nuclei for recrystallization. Treated in this fashion, the activation energy for the recovery process is found to be 19 800 cal mole^{-1}, near to the value of 18 400 cal mole^{-1} obtained for creep. The activation energy for the growth of new material was found to be 31 000 cal mole^{-1} which is near to the value for the activation energy of self diffusion in lead (27 900 cal mole^{-1}).

ACKNOWLEDGMENTS

The above work was carried out at the Royal Institution during the tenure of a Research Fellowship awarded jointly by the Royal Society and the Armourers and Brasiers' Company. I am greatly indebted to all these bodies for their support and for research facilities. The pure lead wire used in the experiments was a gift from the Associated Lead Manufacturers Ltd., who very kindly prepared it specially for the work. Their kindness is gratefully acknowledged.

References

Anderson, W. A., and Mehl, R. F., 1945, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **161,** 140.

Andrade, E. N. da C., 1910, *Proc. Roy. Soc.* A, **84,** 1; 1948, *Nature, Lond.*, **162,** 410.

Andrade, E. N. da C., and Kennedy, A. J., 1951, *Proc. Phys. Soc.* A, **64,** 363.

Avrami, M., 1939, *J. Chem. Phys.*, **7,** 1103; 1940, *Ibid.*, **8,** 212; 1941, *Ibid.*, **9,** 177.

Barrer, R. M., 1941, *Diffusion in and through Solids* (New York: Macmillan).

Burgers, W. A., 1949, *Physica*, **15,** 92.

Burghoff, H. L., and Mathewson, C. H., 1941, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **143,** 45.

Cahn, R. W., 1948, *Strength of Solids* (London: Physical Society), p. 136; 1949, *J. Inst. Metals*, **76,** 121; 1950, *Proc. Phys. Soc.* A, **63,** 323.

Collins, J. A., and Mathewson, C. H., 1940, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **137,** 150.

Cook, M., and Richards, T. L., 1947, *J. Inst. Metals*, **73,** 1.

Cottrell, A. H., 1955, *Proceedings of* 1954 *Symposium on Creep and Fracture of Metals at High Temperatures, Paper No.* 8 (H.M.S.O.), in the press.

Doan, G. E., and Mahla, E. M., 1941, *The Principles of Physical Metallurgy* (New York: McGraw-Hill), p. 126.

Dorn, J. E., 1955, *Proceedings of* 1954 *Symposium on Creep and Fracture of Metals at High Temperatures, Paper No.* 9 (H.M.S.O.), in the press.

Druyvesteyn, M. J., and Manintveld, J. A., 1951, *Nature, Lond.*, **168,** 868.

Eastwood, L. W., Bousu, A. E., and Eddy, C. T., 1935, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **117,** 246.

Gensamer, M., and Mehl, R. F., 1938, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **131,** 372.

Johnson, W. A., and Mehl, R. F., 1939, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **135,** 416.

Karnop, R., and Sachs, G., 1930, *Z. Phys.*, **60,** 464.

Kennedy, A. J., 1952, *Proc. Roy. Soc.* A, **213,** 492; 1953 a, *Nature, Lond.*, **171,** 927; 1953 b, *Brit. J. Appl. Phys.*, **4,** 225; 1953 c, *J. Mech. Phys. Solids*, **1,** 172; 1955, *Nature, Lond.*, in the press.

Kochendörfer, A., 1937, *Z. Kristallogr.*, **92,** 263.

Kornfeld, M. L., and Pavlov, W., 1934, *Phys. Z. Sowjet.*, **6,** 537; 1937, *Ibid.*, **12,** 658.

Kornfeld, M. L., and Schamarin, A., 1937, *Phys. Z. Sowjet.*, **11,** 302.

Manintveld, J. A., 1952, *Nature, Lond.*, **160,** 623.

McKeown, J., 1937, *J. Inst. Metals*, **60,** 201.

Mott, N. F., 1951, *Proc. Phys. Soc.* B, **64,** 729; 1952, *Phil. Mag.*, **43,** 1151; 1953, *Ibid.*, **44,** 742.

Müller, H. G., 1935, *Z. Phys.*, **96,** 279.

Paterson, M. S., and Orowan, E., 1948, *Nature, Lond.*, **162,** 991.

Slifkin, L., and Kauzmann, W., 1952, *J. Appl. Phys.*, **23,** 746.

Smith, A. A., 1941, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **143,** 165.

Smith, A. A., and Howe, H. E., 1945, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **161,** 472.

Smith, A. D. N., 1953, *Phil. Mag.*, **44,** 453.

Sully, A. H., Cale, G. N., and Willoughby, G., 1948, *Nature, Lond.*, **162,** 411.

Wilson, J. E., and Thomassen, L., 1934, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **22,** 769.

Yokobori, T., 1951, *J. Phys. Soc. Japan*, **6,** 404.

A Study of the Thermal Diffusion Equation with Boundary Conditions corresponding to Solidification or Melting of Materials Initially at the Fusion Temperature

By F. KREITH† and F. E. ROMIE‡
† Assistant Professor of Mechanical Engineering, Lehigh University, Bethlehem, Pennsylvania
‡ Department of Engineering, University of California, Los Angeles

Communicated by O. A. Saunders ; MS. received 13th July 1954, and in amended form 30th November 1954

Abstract. Solutions are presented for determining the position of the solid–liquid interface and the temperature–time history of the surface of solidifying or melting materials initially at the fusion temperature. These solutions include treatment of the cylinder, the sphere and the semi-infinite solid. Solutions were obtained by analytical methods and by means of an electrical analogy to the thermal system. The application of the electrical analogy to the solution of problems in solidification is described.

---

§ 1. Introduction

In most heat conduction problems the addition or subtraction of heat from any element serves only to change its temperature. In the case of solidification or melting however the system is essentially more complicated due to the release or absorption of the latent heat of fusion which introduces a non-linearity in the equations describing the process of solidification or melting. Existing solutions to the solidification and melting process have either been limited to particular boundary conditions which eliminate the non-linearity (Stefan 1891, Frank 1927, Ingersoll *et al.* 1948) or have been obtained by postulating that the unit heat capacity of the material is zero (Stefan 1891, London and Seban 1943, Pekeris and Slichter 1939).

This paper presents two new analytical solutions to problems for which the non-linearity is eliminated and two solutions to problems for which the non-linearity is present. The latter two solutions were obtained by means of an electric analogue to the thermal system.

As will be discussed in more detail, a solution to a problem in solidification is also a solution to the corresponding problem in melting. The following development will be arbitrarily restricted to the process of solidification in order to facilitate the presentation.

The solutions presented in this paper are limited to the unidimensional solidification of the semi-infinite solid, the cylinder and the sphere with the latter two shapes solidifying toward the centre. The systems treated are further defined by the following postulates: (i) The material has a definite fusion temperature. (ii) Initially the liquid is uniformly at the fusion temperature. (iii) All thermal properties of the material are uniform and constant. (Properties of the liquid phase however, are not necessarily equal to the corresponding properties of the solid phase.)

In accordance with the foregoing definitions the equations describing the process of solidification are given as follows. Within the solidified phase the temperature is determined by the diffusion equation

$$\nabla^2 T = \frac{1}{a}\frac{\partial T}{\partial \theta}. \qquad \text{......(1)}$$

At the solid–liquid interface the temperature is equal to the fusion temperature. Thus for the semi-infinite solid at $x=\epsilon$ and for the cylinder and sphere at $r=r_0-\epsilon$ (see figures 1 (*a*) and 1 (*b*)) one of the boundary conditions is

$$T = T_f. \qquad \text{......(2)}$$

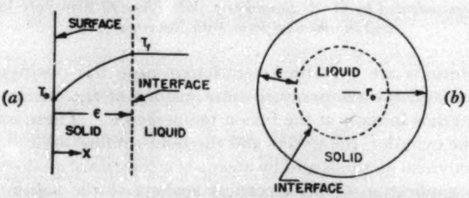


Figure 1. (*a*) Semi-infinite solid; (*b*) cylinder and sphere.

A second condition at the solid–liquid interface is obtained by means of a heat balance which equates the rate at which the heat of fusion is liberated by the solidifying material at the solid–liquid interface to the rate at which the heat is transferred from the interface into the solidified phase. For the semi-infinite solid this heat balance gives

$$k\partial T/\partial x = \rho L d\epsilon/d\theta \qquad \text{at } x=\epsilon \qquad \text{......(3)}$$

where L is the heat of fusion and ρ the mass density of the solidified material. The corresponding equation for the inward solidification of both the cylinder and sphere is

$$k\partial T/\partial r = -\rho L d\epsilon/d\theta \qquad \text{at } r=r_0-\epsilon. \qquad \text{......(4)}$$

The initial conditions are that no solidification has occurred and that the temperature of the liquid is uniformly at the fusion temperature

$$\text{at } \theta=0; \quad \epsilon=0 \quad \text{and } T=T_f. \qquad \text{......(5)}$$

Table 1

Solution number	Shape	Boundary condition Location	Condition	Equation
1	Cylinder	$r=r_0-\epsilon$	$\frac{\partial T}{\partial r}=G$	(6)
2	Sphere	$r=r_0-\epsilon$	$\frac{\partial T}{\partial r}=G$	(7)
3	Semi-infinite solid	$x=\epsilon$	$\frac{\partial T}{\partial x}=G$	(8)
4	,, ,,	$x=0$	$T=T_0$	(9)
5	,, ,,	$x=0$	$k\frac{\partial T}{\partial x}=h(T-T_\infty)$	(10)
6	,, ,,	$x=0$	$\frac{\partial T}{\partial x}=G$	(11)

An additional boundary condition is required before the conditions under which the liquid solidifies are completely specified. Table 1 summarizes the solutions which are presented in this paper by indicating the remaining boundary condition required for complete specification of the problem.

Solutions 3 and 4 have been previously treated by J. Stefan (1890), and Ingersoll *et al.* (1948) and are included in this presentation for completeness and for comparison with solutions to the other systems indicated in table 1.

§ 2. SOLUTION 1

If the temperature gradient in the solidified phase at the solid–liquid interface is specified to be constant as indicated by equation (6) (or by equations (7) and (8)), it then follows from equation (4) that the rate of travel of the interface is constant. Thus it is easily shown from equations (4) and (6) that the position of the interface at any time θ after the initiation of solidification is given, for the solidification of the cylinder, by the equation

$$\epsilon/r_0 = -kG\theta/\rho L r_0 \qquad \ldots\ldots(12)$$

where the outer radius r_0 of the cylinder has been introduced in order to express the relation in a convenient dimensionless form. The dimensionless modulus ϵ/r_0 may be noted to vary from an initial value of zero to unity when the cylinder is completely solidified.

In order to determine the temperature distribution in the solidified phase at any time, and, more particularly, the temperature–time history which must be imposed on the outer surface of the cylinder in order to maintain a constant rate of solidification, the solution to equations (1), (2), (4), (5) and (6) must be obtained. These equations may be written in the convenient dimensionless form

$$\left.\begin{aligned} &\frac{1}{r^+}\frac{\partial}{\partial r^+}\left(r^+\frac{\partial T^+}{\partial r^+}\right) = G^+\frac{\partial T^+}{\partial \theta^+}\,; \\ \text{at}\quad & r^+ = 1-\epsilon^+, \qquad T^+ = 0, \qquad \partial T^+/\partial r^+ = -d\epsilon^+/d\theta^+ = +1 \\ \text{at}\quad & \theta^+ = 0, \qquad T^+ = 0, \qquad \epsilon^+ = 0 \end{aligned}\right\} \quad \ldots\ldots(13)$$

upon defining the following dimensionless moduli:

$$T^+ = \frac{T-T_f}{r_0 G}, \qquad r^+ = \frac{r}{r_0}, \qquad \theta^+ = \frac{\theta k G}{\rho L r_0}, \qquad \epsilon^+ = \frac{\epsilon}{r_0}, \qquad G^+ = \frac{Gcr_0}{L}$$

where T_f is the temperature of fusion and c is unit heat capacity.

The solution to equations (13) may be obtained by means of a method of iterative approximation in which the temperature modulus T^+ is expressed in the series:

$$T^+ = T_1 + T_2 + T_3 + \ldots.\ T_n + \ldots. \qquad \ldots\ldots(14)$$

Each term of this series may be related to the preceding term by the equation

$$\frac{1}{r^+}\frac{\partial}{\partial r^+}\left(r^+\frac{\partial T_n}{\partial r^+}\right) = G^+\frac{\partial T_n - 1}{\partial \phi} \qquad \ldots\ldots(15)$$

where, by definition, $\phi = 1 + \theta^+$.

After integration this equation becomes

$$T_n = G^+ \int_{\phi}^{r^+} \frac{dr^+}{r^+} \int_{\phi}^{r^+} r^+ \frac{\partial T_n - 1}{\partial \phi}\, dr^+ \qquad (n>1). \qquad \ldots\ldots(16)$$

The first term, T_1 of the series is obtained by setting $G^+=0$ in equation (15). The first five terms of the series are presented below. (The superscript + is omitted from the moduli G^+ and r^+.)

$$\left.\begin{aligned}
T_1 &= \phi \ln \frac{r}{\phi} \\
T_2 &= -G\left[\left(\frac{r^2}{4} + \frac{3}{4}\phi^2\right)\ln\frac{r}{\phi} + \frac{\phi^2}{2} - \frac{r^2}{2}\right] \\
T_3 &= G^2\left[\left(\frac{5}{16}\phi^3 + \frac{3}{8}\phi r^2\right)\ln\frac{r}{\phi} + \frac{21}{64}\phi^3 - \frac{5}{16}\phi r^2 - \frac{r^4}{64\phi}\right] \\
T_4 &= -G^3\left[\left(\frac{35}{384}\phi^4 + \frac{15}{64}\phi^2 r^2 + \frac{3}{128}r^4\right)\ln\frac{r}{\phi}\right. \\
&\qquad \left. + \frac{278}{2304}\phi^4 - \frac{17}{256}\phi^2 r^2 - \frac{7}{128}r^4 + \frac{r^6}{2304}\phi^2\right] \\
T &= G^4\left[\left(\frac{198}{9216}\phi^5 + \frac{35}{384}\phi^3 r^2 + \frac{15}{512}r^4\phi\right)\ln\frac{r}{\phi}\right. \\
&\qquad + \frac{2325}{73728}\phi^5 + \frac{31}{4608}\phi^3 r^2 - \frac{77}{2048}\phi r^4 \\
&\qquad \left. - \frac{1}{1536}\frac{r^6}{\phi} - \frac{1}{73728}\frac{r^8}{\phi}\right].
\end{aligned}\right\} \qquad \ldots\ldots(17)$$

For values of the modulus G^+ less than about $-1{\cdot}5$ the series converges very slowly with the result that the first five terms given above are insufficient for an accurate evaluation of T^+ when G^+ is less than about $-1{\cdot}5$. For values of G^+ equal to or greater than $-1{\cdot}5$ the evaluation of T^+ by means of the first five terms given is estimated to be accurate to within one or two per cent of the exact value.

The temperature–time history which must be imposed on the surface of the cylinder in order to maintain a constant rate of travel of the interface has been evaluated from the series solution and is presented in figure 2 for values of the modulus G^+ from zero to $-1{\cdot}5$. Discussion of figure 2 will be deferred until the solidification of the sphere has been presented.

§ 3. SOLUTION 2

If the sphere is solidified in such a manner that the temperature gradient in the solidified phase at the solid–liquid interface is maintained constant, it follows, from equations (4) and (7), that the position of the interface at any time is given by the equation

$$\frac{\epsilon}{r_0} = -\frac{kG\theta}{\rho L r_0} \qquad \ldots\ldots(18)$$

which is identical with equation (12) given previously for the cylinder.

In order to determine the temperature distribution in the solidified portion of the sphere at any time the solution to equations (1), (2), (4), (5) and (7) must be obtained. Employing the same dimensionless parameters defined for the cylinder the equations for the sphere can be expressed in the form:

$$\left.\begin{array}{l} \dfrac{1}{r^{+2}}\dfrac{\partial}{\partial r^{+}}\left(r^{+2}\dfrac{\partial T^{+}}{\partial r^{+}}\right)=G^{+}\dfrac{\partial T^{+}}{\partial \phi} \quad \text{at} \quad r^{+}=1-\epsilon^{+} \\ \dfrac{\partial T^{+}}{\partial r^{+}}=-\dfrac{d\epsilon^{+}}{d\theta^{+}}=+1, \quad T^{+}=0 \quad \text{at} \quad \theta^{+}=0, \quad \epsilon^{+}=0, \quad T^{+}=0. \end{array}\right\} \quad \ldots\ldots(19)$$

The solution to these equations may be obtained in a manner exactly analogous to that used for the cylinder. Thus, expressing the solution in series form

$$T^{+}=T_1+T_2+T_3+\ldots T_n+\ldots \qquad \ldots\ldots(20)$$

any term of the series may be related to the preceding term by the equation

$$\frac{1}{r^{+2}}\frac{\partial}{\partial r^{+}}\left(r^{+2}\frac{\partial T_n}{\partial r^{+}}\right)=G^{+}\frac{\partial T_{n-1}}{\partial \phi} \qquad \ldots\ldots(21)$$

which gives, on integration

$$T_n=G^{+}\int_{\phi}^{r^{+}}\frac{dr^{+}}{r^{+2}}\int_{\phi}^{r^{+}} r^{+2}\frac{\partial T_{n-1}}{\partial \phi}\,dr^{+} \quad (n>1). \qquad \ldots\ldots(22)$$

As was done for the cylinder the first term T_1 of the series is obtained by setting $G^{+}=0$ in equation (21). Upon evaluation of the first several terms of the series it is possible, by inspection, to determine the law by which the series progresses. The solution obtained in this manner, which may be shown to be an exact solution of equations (19), is

$$T^{+}=-\sum_{n=1}^{\infty} G^{n}r^{n+1}\frac{n}{(n+2)!}+\sum_{i=-1}^{+\infty}\sum_{j=0}^{\infty}(-1)^{i-1}G^{i+j}r^{j-1}\phi^{i+2}\frac{2i+j+2}{(i+2)!\,j!}. \quad \ldots(23)$$

(In this equation the superscript + has been omitted from the moduli G^{+} and r^{+}.)

The temperature–time history which must be imposed on the outer surface of the sphere in order to maintain a constant rate of travel of the solid–liquid interface has been evaluated from equation (21) and is presented in figure (3) for values of the modulus G varying from zero to $-1{\cdot}5$.

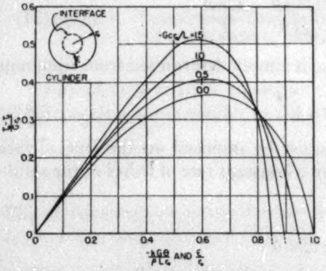


Figure 2. Temperature–time history which must be imposed on the surface of a cylinder solidifying at a constant rate.

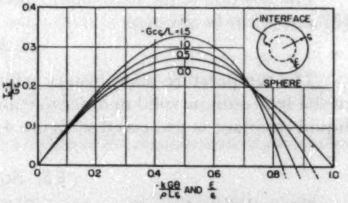


Figure 3. Temperature–time history which must be imposed on the surface of a sphere solidifying at a constant rate.

Inspection of both figure 2 for the cylinder and figure 3 for the sphere shows that it is not possible to solidify either shape in such a manner that the velocity of the solid–liquid interface is maintained constant during the final stages of solidification. Near the end of solidification, as the solid–liquid interface approaches the centre of these two shapes, the surface temperatures required by the solutions to cases 2 and 3 of table 1 exceed the fusion temperature of the material. In other words, to maintain a constant rate of travel of the interface it would be necessary to raise the temperature at the outer surface above the melting point. This would cause the surface layer to melt, a condition not provided for in the mathematical formulation of the problem. Hence, the solutions as given by equations (17) and (23) apply only until θ^+ and ϵ^+ reach the value at which the curves of T^+ cross the abscissae of figures 2 or 3 (for instance, at $G^+ = -1{\cdot}0$ for the cylinder the solution is valid until $\theta^+ = 0{\cdot}89$ i.e. until the frozen interface reaches a distance $0{\cdot}11r_0$ from the centre). If the surface temperature would be maintained from that time on at the fusion temperature, the interface would advance at an increased rate. This rather curious result becomes apparent when one observes that in the last stages of a constant rate solidification of a cylinder or a sphere (i.e. beyond the maximum of the T^+ curves) the frozen material between the solidifying interface and the surface acts as a heat sink having a capacity exceeding the requirements. Therefore, to maintain a constant rate of advance of the interface as the flow rate of latent heat from the centre decreases, it is necessary to increase to surface temperature in order to reduce the temperature gradient at the interface. Once the surface temperature reaches the melting point and is maintained there, the temperature gradient at the solidifying interface will increase and, as stated previously, cause the centre portion to freeze at an increased rate.

§ 4. SOLUTION 3

The solidification of a semi-infinite solid in such a manner that the solid–liquid interface proceeds at a constant rate has been previously treated by Stefan (1891). This mode of solidification is described by equations (1), (2), (3), (5) and (8). The solution to these equations gives for the temperature distribution in the solidified phase

$$\frac{(T-T_f)c}{L} = 1 - \exp\left(\frac{G^2ck\theta}{\rho L^2} - \frac{Gcx}{L}\right). \qquad \ldots\ldots(24)$$

The position of the solid–liquid interface is simply determined from equations (3) and (8) to be given as

$$\epsilon\rho L/kG\theta = 1. \qquad \ldots\ldots(25)$$

The temperature–time history which must be imposed on the free surface of the semi-infinite solid in order to maintain a constant rate of travel of the solid–liquid interface is presented in figure 4.

§ 5. SOLUTION 4

The solidification of a semi-infinite body of liquid initially at a uniform temperature above fusion was first solved analytically by Franz Neumann (Frank 1927, Ingersoll *et al.* 1948) for the case in which the surface temperature was suddenly depressed to and maintained at a constant temperature T_0 below the fusion temperature. This solution was later simplified by Stefan (1891), and Ingersoll *et al.* (1948) to the case in which the liquid was initially uniformly at the fusion temperature. This latter process of solidification is described by equations (1), (2), (3), (5) and (9) which give upon solution

$$\frac{T-T_f}{T_0-T_f}=1-\frac{\operatorname{erf} x/2\sqrt{(a\theta)}}{\operatorname{eif}\beta} \qquad \ldots\ldots(26)$$

for the temperature distribution within the solidified phase. The position of the solid–liquid interface is expressed by the equation

$$\epsilon^2/a\theta=4\beta^2. \qquad \ldots\ldots(27)$$

In these two equations the parameter β is an eigenvalue which is determined from the eigenfunction

$$\sqrt{\pi}\beta\exp(\beta^2)\operatorname{erf}\beta=\frac{(T_f-T_0)c}{L}. \qquad \ldots\ldots(28)$$

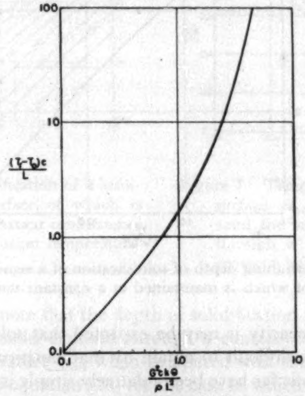


Figure 4. Temperature–time history which must be imposed on the surface of a semi-infinite solid solidifying at a constant rate.

From an inspection of equations (27) and (28) it is apparent that the position of the interface is expressible in the form

$$\frac{\epsilon^2}{a\theta}=f\left(\frac{L}{(T_f-T_0)c}\right). \qquad \ldots\ldots(29)$$

This relationship is presented in figure 5 for values of the dimensionless parameter $L/(T_f-T_0)c$ ranging from 0·10 to 100.

§ 6. Solution 5

The solidification of a semi-infinite body of liquid initially at the fusion temperature is described by equations (1), (2), (3), (5) and (10) for the case in which the free surface is exposed through a constant and uniform thermal conductance to a fixed temperature below the fusion temperature of the liquid.

Before considering the solution to these equations it may be observed that one of the boundary conditions for the diffusion equation is non-linear in T. Specifically, since T is a function of both time and position, one may write:

$$dT = \frac{\partial T}{\partial x}\,dx + \frac{\partial T}{\partial \theta}\,d\theta. \qquad \ldots\ldots(30)$$

Employing equations (2) and (3) it follows from equation (30) that one of the boundary conditions at the solid-liquid interface may be expressed as

$$\frac{\partial T}{\partial \theta} = -\frac{k}{\rho L}\left(\frac{\partial T}{\partial x}\right)^2. \qquad \ldots\ldots(31)$$

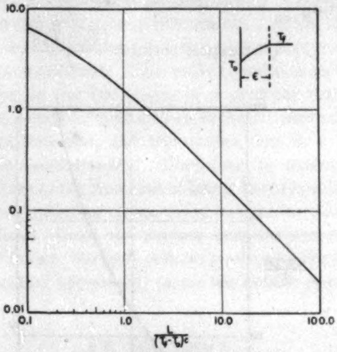


Figure 5. Curve for determining depth of solidification of a semi-infinite solid the surface temperature of which is maintained at a constant temperature T_0.

Due to this non-linearity it may be expected that solutions to the problem of solidification will be difficult to obtain by mathematical methods. That the solutions considered thus far have been relatively simple is due, in the case of the first three solutions, to the fact that the temperature gradient at the interface was stated to be a constant, which naturally eliminated the non-linearity involved in equation (31). In the case of the fourth solution the depth ϵ of solidified material was found to increase proportionately with the square root of time. This relationship, as may easily be verified, also served to make equation (31) linear.

The solution to the solidification of a semi-infinite solid with its face exposed through a thermal resistance to a below-fusion temperature was, in view of the difficulties implied by the boundary equation (31), not obtained by mathematical methods but by means of a thermal analyser. The thermal analyser is an analogue device by means of which the thermal system is simulated by an electrical network consisting of electrical resistors and capacitors. A description of the thermal analyser and an account of the methods employed in obtaining solutions to problems in solidification is given subsequently.

A dimensional analysis of equations (1), (2), (3), (5) and (10) shows that the position of the interface may be expressed as a function of two parameters:

$$\frac{\epsilon h}{k} = f_1\left[\frac{L}{(T_f - T_\infty)c}, \frac{ah^2\theta}{k^2}\right] \qquad \ldots\ldots(32)$$

and that the temperature within the solidified portion is a function of three dimensionless parameters:

$$\frac{T-T_\infty}{T_f-T_\infty}=f_2\left[\frac{L}{(T_f-T_\infty)c},\ \frac{ah^2\theta}{k^2},\ \frac{hx}{k}\right]. \qquad \ldots\ldots(33)$$

The position of the solid–liquid interface, as obtained by means of the thermal analyser, is presented in figure 6. The temperature–time history of the surface ($hx/k=0$) of the semi-infinite solid is presented in figure 7. These solutions, obtained on the thermal analyser are estimated, by methods presented in the description of the thermal analyser, to differ by less than three per cent from the exact solution to the differential equations.

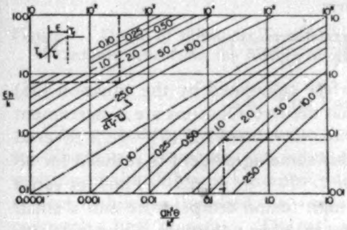


Figure 6. Depth of solidification of a semi-infinite solid the surface of which is exposed through a constant conductance to a constant below-fusion temperature.

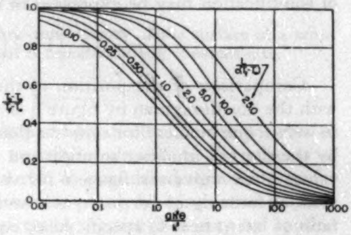


Figure 7. Temperature–time history of the surface of a solidifying semi-infinite solid the surface of which is exposed through a constant conductance to a constant below-fusion temperature.

It is of interest to note that the depth of solidification predicted from figure 6 for the semi-infinite solid exposed through a constant surface conductance to a below-fusion temperature T_∞, when compared with the depth of solidification predicted by means of equation (26) for the case in which the surface is suddenly depressed to and maintained at the same temperature T_∞, shows that the two depths of solidification differ by less than ten per cent provided the modulus $\epsilon h/k$ is greater than unity. The agreement of the two predicted depths of solidification improves as $\epsilon h/k$ becomes larger, being better than the scatter of data for the thermal analyser solution when the modulus $\epsilon h/k$ becomes equal to or greater than 10.

The depth of solidification predicted by the thermal anlyser solution will always be somewhat less than that predicted by equation (26) due to the thermal resistance on the surface of the semi-infinite solid in the former case. The agreement between the two solutions in predicting the depth of solidification at large values of $\epsilon h/k$ is easily explained by the following reasoning. The heat transferred from the travelling interface 'sees' two thermal resistances in series; the resistance to flow from the interface to the surface of the semi-infinite solid and the resistance (represented by the reciprocal of the unit thermal conductance h) to flow from this surface to the surrounding medium. (For a fixed surface temperature this latter resistance is zero.) Thus when the depth of solidification becomes sufficiently large the resistance to flow of heat from the surface becomes negligible in comparison with the resistance of the solidified material and, consequently, the solidification rate will proceed as though there were no surface resistance.

During the initial stages of solidification the thermal capacity of the solidified phase has little effect on the solidification rate of the material particularly if the modulus $L/(T_f - T_\infty)c$ is large. Thus a first approximation to the position of the interface is obtained by setting the unit heat capacity equal to zero. The equations describing the solidification of the solid are then equations (2), (3), (5) and (10) with equation (1) replaced by the equation

$$\partial^2 T/\partial x^2 = 0. \qquad \ldots\ldots(34)$$

The solution (London and Sebian 1943) to this set of equations for the depth of solidification may be expressed in the form

$$\frac{\epsilon h}{k} = -1 + \left\{1 + 2\left(\frac{ah^2\theta}{k^2}\right)\left(\frac{c(T - T_\infty)}{L}\right)\right\}^{1/2}. \qquad \ldots\ldots(35)$$

Comparison of the position of the interface calculated by the equation (35) with the position given by figure 6 shows that the two solutions are in agreement to within one per cent for $\epsilon h/k$ less than 0·01 for all values of $L/(T_f - T_\infty)c$ covered by the thermal analyser solution and that this same agreement is realized for all values of $\epsilon h/k$ given in figure 6 for values of $L/(T_f - T_\infty)c \geqslant 25$. Thus for rapid rates of freezing or for materials having a high fusion temperature and a small ratio of latent heat to specific heat, equation (35) does not apply. For instance, at $L^+ = 0{\cdot}1$ the depth of freezing predicted from equation (35) at $\theta^+ = 1{\cdot}0$ is twice as large as the value of ϵ^+ determined by means of the thermal analyser. Figure 6 should thus be used for metals such as lead, lithium, magnesium and aluminium as well as for refractory materials such as are used in certain rocket motors and other high temperature devices.

From the foregoing discussion it is concluded that the solution to equations (1), (2), (3), (5) and (10) for the position of the interface is available for any depth of solidification provided that $L/(T_f - T_\infty)c \geqslant 0{\cdot}10$.

§ 7. SOLUTION 6

The solidification of a semi-infinite solid when the rate of heat transfer from the surface is maintained constant (i.e. when the temperature gradient at the surface is maintained constant) is described by equations (1), (2), (3), (5) and (11). When these equations are written employing dimensionless parameters it is easily seen that the position of the interface and the temperature of the surface of the semi-infinite solid are functions of a single variable. Thus

$$\frac{\epsilon Gc}{L} = f_1\left(\frac{aG^2c^2\theta}{L^2}\right) \qquad \ldots\ldots(36)$$

$$\frac{(T_f - T_0)c}{L} = f_2\left(\frac{aG^2c^2\theta}{L^2}\right). \qquad \ldots\ldots(37)$$

The solutions represented by these two equations have been obtained by means of the thermal analyser and are presented in figure 8. A partial check on the accuracy of the solutions was made by obtaining the solution to the equations by means of finite difference calculations. The difference solution, which was tedious to obtain, was not carried out for a depth of solidification greater than $\epsilon Gc/L = 3{\cdot}0$, but within this depth the thermal analyser and difference technique solutions differed by less than 5%. It is estimated that the solutions represented by figure 8 differ by less than 5% from the exact solution to equations (1), (2), (3), (5) and (11).

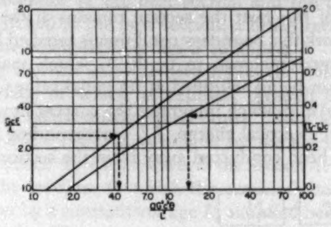


Figure 8. Depth of solidification and temperature–time history of the surface of a semi-infinite solid from the surface of which heat is transferred at a constant rate.

§ 8. The Process of Melting

The preceding solutions have been obtained specifically for the process of solidification. However, the solutions are equally applicable to the process of melting of the different systems considered if it is postulated that the transfer of heat within the liquefied (i.e. melted) material is purely by conduction without convective heat transfer. The solutions given by figures 2–6 may be applied directly to the process of melting by the simple artifice of defining the heat of fusion to be a negative number for melting but a positive number for solidification. For melting, the thermal properties, k, c, a and ρ are naturally evaluated for the liquid phase.

§ 9. Method of Analogue Solution

The thermal analyser (Kreith 1949, Paschkis and Baker 1942) is an analogue computer in which thermal capacity and resistance are simulated by electrical capacities and resistances and in which heat flow and temperature differences are analogous to electrical current and potential differences respectively. In the application of the analyser the thermal system (e.g. the solidified phase of a solidifying semi-infinite solid) is divided by surfaces parallel to the exposed surface into sections of thickness Δx. Each section has a thermal capacity and thermal resistance, both of which are treated as lumped parameters. This system of lumped thermal resistances and capacities is then simulated in the analyser by a network of lumped electrical resistances and capacities. For the case of constant thermal properties the analogy existing between the lumped thermal circuit and the electrical network is exact.

Figure 9 shows the electrical network and associated components used to solve the problem of the solidification of a semi-infinite solid initially at the fusion temperature. For this problem the thermal system has been divided into sections of equal thickness so that the thermal capacities, C and thermal resistances R (and consequently the analogous electrical capacities C' and resistance R') are the same for each section. The thermal resistance of a section of thickness x and area normal to heat flow A is $R=\Delta x/kA$. The thermal capacity of the section is $C=\Delta xA\rho c$. Approximately 45 resistance–capacity sections were employed for the solution of the solidification problems presented.

Before any one section of the lumped thermal system can be solidified the heat of fusion of the section must be conducted into the solidified portion of the semi-infinite solid. The heat of fusion L_s for a section is $L\rho\Delta xA$. Previous to and during the removal of this heat the section remains at the fusion temperature. In the electrical network this boundary condition is satisfied by use of an electrical component (the integrator shown in figure 9) which maintains the electrical potential of the section being solidified at its initial value (corresponding to the temperature of fusion) and which integrates the current flowing from the section until a predetermined electrical charge, L_s', corresponding to the heat of fusion of the section L_s has been conducted away from the section.

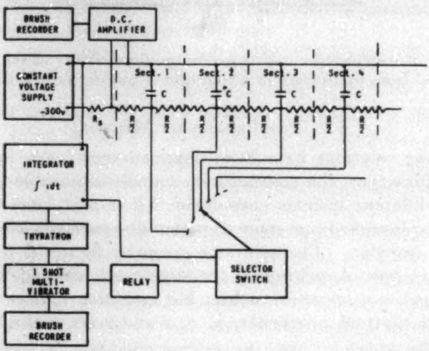

BRUSH RECORDER
D.C. AMPLIFIER
CONSTANT VOLTAGE SUPPLY
-300v
Sect. 1
Sect. 2
Sect. 3
Sect. 4
C
R_s
R/2
INTEGRATOR
∫ idt
THYRATRON
1 SHOT MULTI-VIBRATOR
RELAY
SELECTOR SWITCH
BRUSH RECORDER

Figure 9. Schematic diagram of the electrical components of the thermal analyser.

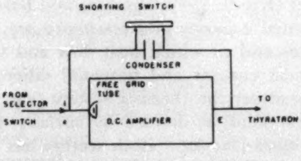

SHORTING SWITCH
CONDENSER
FREE GRID TUBE
FROM SELECTOR SWITCH
i
e
D.C. AMPLIFIER
E
TO THYRATRON

Figure 10. Schematic diagram of the $\int idt$ integrator.

The integrator, shown schematically in figure 10, consists essentially of a d.c. amplifier across which a condenser is shunted. The amplifier has a gain which produces a large positive voltage E for a small drop in voltage e on the input side. When a current i flows from the selector switch towards the amplifier and produces a small negative voltage this voltage is amplified through a free-grid electronic tube, and a large positive voltage is imposed on the condenser. The condenser then absorbs the current flowing towards the amplifier from the network while at the same time a constant voltage is maintained at the particular section in the process of 'freezing'. It can easily be shown that for the type of electronic network described above the rise of voltage E is proportional to the time integral of the current, $\int idt$, which is the electrical quantity analogous to the latent heat of fusion.

When the voltage E has reached a certain predetermined value (i.e. the charge L_s' has been transferred from the section) the integrator energizes a solenoid which switches the integrator to the next section and the process is repeated. As the integrator is switched to the next section, a brush recorder records a pip which indicates the time when each section is solidified.

Figure 9 shows the thermal analyser network for the solidification of a semi-infinite solid, the surface of which is exposed through a constant thermal resistance R_s to a fixed below-fusion temperature. In the electrical network this thermal resistance ($R_s = 1/hA$) is simulated by the electrical resistance R_s'. The temperature difference $T_f - T_\infty$ between the fusion temperature and the fixed temperature of the medium to which the semi-infinite solid is exposed is simulated in the analyser by a constant voltage E_0 supplied by the constant voltage source shown in figure 9. The temperature–time history of the surface of the slab is determined from the tape of the brush recorder which is shown connected through a high impedance amplifier to the 'surface' of the semi-infinite solid.

Translation of data obtained on the thermal analyser into thermal terms is most easily accomplished by noting that, due to the exact analogy existing between the lumped thermal and the electrical systems, dimensionless moduli in the electrical system must be numerically equal to the corresponding moduli in the thermal system. Thus, employing relations previously given,

$$\frac{L}{(T_f - T_\infty)c} = \frac{L_s}{(T_f - T_\infty)c} = \frac{L_s'}{c'E_0}$$

$$\frac{ah^2\theta}{k^2} = \frac{R\theta}{R_s{}^2 c} = \frac{R'\theta'}{R_s'^2 c'}$$

$$\frac{\Delta\epsilon h}{k} = \frac{\Delta x h}{k} = \frac{R}{R_s} = \frac{R'}{R_s'}$$

and

$$\frac{T - T_\infty}{T_f - T_\infty} = \frac{E}{E_0}.$$

To span the ranges of variables of practical importance a survey of materials and operating conditions of interest in engineering and geology was made prior to the construction of the thermal analyser. For a thorough coverage of the variables it was found necessary to build each section with two variable resistors which were logarithmically wound and covered a range of resistances from 10^3 to 10^6 ohms and provide each section with six oil-filled paper condensers of 0·25, 0·5, 1·0, 2·0, 4·0 and 8·0 microfarads capacity.

The accuracy of the solutions obtained on the thermal analyser was estimated by solving on the thermal analyser the problem of the semi-infinite solid with an infinite surface conductance (i.e. no electrical resistance R_s'). Since this boundary condition corresponds to that of solution 4 for which an exact analytical solution is available, the accuracy of the thermal analyser solution could be estimated by comparing the analyser and analytic solutions. By this method the solutions obtained by means of the analyser are estimated to differ from the exact solutions by less than three per cent.

It is of interest to note that solutions 5 and 6, which were quite simply obtained by means of the thermal analyser, were not obtained by analytical methods due to the mathematical difficulties implied by the non-linear boundary conditions.

In contrast, solutions 1, 2 and 3, which were quite simply obtained by analytical methods, could have been obtained on the thermal analyser only by a trial and error method due to the form in which the boundary conditions occur. In particular, attainment of solutions 1, 2 and 3 by means of the thermal analyser would have entailed the assumption of a surface temperature–time history, application of the analogue of this history to the analyser, and a comparison of the resultant travel of the liquid–solid interface with the travel required by the boundary condition. This process would then have to be repeated until the surface temperature–time history consistent with the requirement of constant rate of interface travel was attained. Thus, for the solutions presented in this paper the two methods of solution were complementary.

Table 2. Survey of Physical Properties for Typical Materials of Interest in Melting and Freezing†

Name	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Aluminium	27	168	168	150	0·26	1200	3·44	646
Ammonia	17		50	0·3	1·0	−122	0·015	
Aniline	93	198	63	0·1	0·46	21	0·035	430
Benzene	78	172	44	0·1	0·42	43	0·054	410
Beryllium	9	622	106	115	0·43	2370	2·52	1450
Bismuth	209	22·4	605	4	0·03	520	0·22	747
Copper	64	78	555	200	0·11	1980	3·28	709
Lead	207	9·8	700	18	0·033	621	0·78	297
Lithium	10	59	33	25	0·80	360	0·95	73·8
Magnesium	24	126	106	92	0·25	1200	3·45	504
Mercury	200	5·0	840	5	0·03	−38	0·02	167
Nickel	59	133	555	30	0·14	2640	0·38	950
Sodium	23	49·5	63	80	0·3	208	4·23	165
Steel	56	117	485	20	0·14	2790	0·29	836
Tin	119	26	460	30	0·05	450	1·30	520
Tungsten	184	79	1180	115	0·04	~6000	2·43	1975
Water	18	144	63	0·3	1·0	32	0·048	144

† The following references were consulted :

Heat Transmission by W. H. McAdams, 2nd Edn, 1942, New York : McGraw-Hill.
Chemical Engineers' Handbook by John H. Perry, 2nd Edn, 1941, New York : McGraw-Hill.
Metals Handbook, 1939, Ed. American Society for Metals.
Mechanical Properties of Metals and Alloys, Circular C 447. Dec. 1943, U.S. Government Printing Office, Washington, D.C.
Mechanical Engineers' Handbook by L. S. Marks, 4th Edn, 1941, New York : McGraw-Hill.

(1) Molecular weight; (2) heat of fusion L (BTU lb^{-1}); (3) density ρ (lb ft^{-3}); (4) thermal conductivity k (BTU ft^{-1} deg^{-1} h^{-1}); (5) specific heat (BTU lb^{-1} deg^{-1}); (6) fusion temperature (°F); (7) thermal diffusivity (ft^2 h^{-1}); (8) L/c (°F).

In order to facilitate the application of the results of this study in practice, the physical properties of a variety of materials commonly met in engineering are given in table 2. From this tabulation the pertinent values of L^+, θ^+, T^+ and G^+ can readily be determined and by means of the charts of this paper applied to specific problems.

The thermal analyser is suitable and easily adaptable to the solution of other one-dimensional problems in solidification such as the solidification of a cylinder or a sphere. It is also possible to use the analogue for different boundary conditions. For example, a variation with time of the thermal conductance at the surface of the body can be simulated on the analogue network by a corresponding variation of the electrical resistance R_s' in figure 9. A variation of the temperature of the surrounding medium can be simulated by a corresponding variation of the voltage potential across the electrical network. A specified rate of heat flow can be simulated by replacing the voltage supply shown in figure 9 by a current supply. A superheated melt with convection heat transfer from the melt to the interface can be simulated by inserting appropriate resistances in each of the electrical leads from individual sections to the selector switch contact points shown in figure 9. Also the solution of two dimensional problems is feasible although a separate integrator would be required for each lumped section.

REFERENCES

FRANK, P., 1927, *Die Differential- und Integralgleichung der Mechanik und Physik*, 7th Edn, Vol. 2 of Riemann–Weber (Braunschweig: Vieweg und Sohn), pp. 220–222.

KREITH, F., 1949, *Thesis*, University of California.

INGERSOLL, L. R., ZOBEL, O. J., and INGERSOLL, A. C., 1948, *Heat Conduction with Engineering and Geological Applications*, 1st Edn (New York: McGraw-Hill).

LONDON, A. L., and SEBAN, R. A., 1943, *Trans. Amer. Soc. Mech. Engrs*, **65**, 7.

PASCHKIS, V., and BAKER, H. D., 1942, *Trans. Amer. Soc. Mech. Engrs*, **64**, 105.

PEKERIS, C. L., and SLICHTER, L. B., 1939, *J. Appl. Phys.*, **10**, 135.

STEFAN, J., 1891, *Ann. Phys. Chem.*, **42**, 269.

Noise from Hot-cathode Discharges

By THONG SAW PAK†
Physics Department, Queen's University, Belfast

Communicated by K. G. Emeléus; MS. received 13*th November* 1954

Abstract. A study has been made of the low-frequency noise which occurs in hot cathode discharges through mercury vapour at low pressure. Under non-saturated conditions the whole plasma fluctuates in potential coherently, probably on account of fluctuations of the cathode sheath. The effects of these fluctuations on probes, used under different conditions, are discussed. There is evidence for the existence of a second type of noise, associated more directly with fluctuations of ion sheaths on the wall or probes.

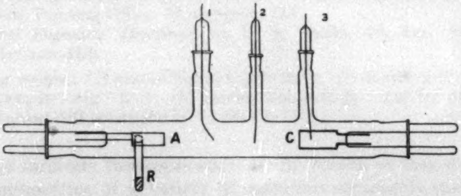
---

§ 1. Introduction

Low frequency noise from low pressure gas discharges has been investigated by Cobine and Gallagher (1947), Martin and Woods (1952), and others. The noise is mainly ' white ', with a frequency spectrum extending up to about 1 Mc/s, above which there is a rapid decrease in amplitude. The transition frequency is related to the ion mass of the gas, and the noise amplitude integrated over all frequencies can be as large as two volts. The noise is not related to plasma volume and is quite different from shot or thermal noise. The present paper deals mainly with the speed and mode of propagation of the noise in the plasma and the manner in which it affects internal and external probes. The results indicate that it is closely connected with the fluctuation of space-charge sheaths.

§ 2. Experimental Arrangements

The experimental tube is shown in the figure. It was connected to a mercury diffusion pump backed by a rotary oil pump, and contained mercury vapour at



Discharge tube. A anode, C cathode, R mercury reservoir, 1, 2, 3 leads to probes, the latter not shown in detail. In some experiments the filament of C was to the right of probe 3.

---

† Now at Physics Department, University of Malaya, Singapore.

pressures between 7×10^{-4} and 2×10^{-3} mm Hg. The cathode was a 200 micron tungsten filament 4 cm in length, and the anode a hollow cylinder of thin nickel sheet 2·5 cm in diameter and 6 cm long. The tube current saturated approximately at about 30 v between anode and cathode. Maximum tube currents of about 300 mA were passed. The probes were squares of molybdenum $0{\cdot}2 \times 0{\cdot}2$ cm^2, 0·015 mm thick, or fine wires, supported at the ends of glass-shielded thin nickel or tungsten wires. All electrodes were mounted on standard ground joints. The positions of the probes were readily altered by having their leads skew, as indicated, and rotating the joints holding them.

Accumulators were used for the h.t. and l.t. supplies, in conventional tube and probe circuits. Probe circuits were connected to the anode, which was earthed. Two identical filter systems were used in part of this work. In each, the filter was preceded by a cathode follower and an amplifier. Both the cathode-follower and amplifier circuits were conventional. The filter was based on a circuit described by Terman *et al.* (1939) ; a frequency range of 60 kc/s to 3 Mc/s was covered and narrow bandwidths were used, e.g. 300 c/s at the 3 dB point for a centre frequency of 70 kc/s.

The simplest way of viewing and measuring the integrated noise was to use an oscilloscope with an amplifier response covering the required frequency range. A Furzehill oscilloscope, type 1684 D/2 was used for this purpose. Its amplifier response is uniform from d.c. to 1 Mc/s. Since the noise amplitude for a mercury vapour discharge above 1 Mc/s is low and rapidly decreasing with frequency, the falling amplification characteristic above 1 Mc/s will have a negligible effect.

When a quantitative measure of the noise for different discharge parameters was needed, an estimate of the mean noise amplitude, obtained over the full bandwidth, was made over several superposed sweeps of the repetition time-base of the oscilloscope. This mean value was treated as if it had been a sine-wave of the same amplitude, and the integrated noise then calculated in decibels above a convenient reference level, viz. 9 mv peak to peak.

The experiments were done in an electrostatically shielded cage.

§ 3. EXPERIMENTAL RESULTS

With the cylindrical anode, anode spots, which are very unstable at a disc anode, were either absent or appeared inside the cylinder. The discharge was then stable and the tube filled with plasma. The low frequency noise persisted, however. The results of Martin and Woods (1952) concerning the relationship between tube noise and the tube characteristic curve were confirmed for the integrated tube noise. In particular, throughout the non-saturated part of the characteristic curve, the noise was found to be high. It dropped to a low level with current saturation.

An attempt was made to measure the velocity of propagation of tube noise, using the probes two at a time, with each connected to a voltage bias system. One probe was connected to an X and the other to a Y plate of the oscilloscope, through the identical filter systems, and the two remaining plates of the oscilloscope were earthed. In this manner a narrow frequency band of the noise was selected from two different parts of the plasma to form an ellipse on the oscilloscope. A change in phase shown by the ellipse with change in probe position would indicate the presence of a travelling wave. It was found, under non-saturated conditions, that the fluctuations constituting the noise occurred simultaneously throughout the tube, within the limits of accuracy of the experiment. For example, the speed of the noise component with frequency 70 kc/s could not have been less than 10^7 cm sec^{-1}. This is several orders of magnitude greater than the speed of positive ion plasma waves of small amplitude (Tonks and Langmuir 1929) under the conditions of ionization in the tube, and shows that the noise is not associated with these oscillations.

Ionic space-charge sheaths at the probes both attenuated the probe current fluctuations and introduced a phase change. This is probably due to their capacitive reactance, which has been studied more directly by Appleton and Childs (1930).

In a non-saturated discharge a sharp maximum of noise occurs at a probe at space potential (Armstrong, Emeléus and Neill 1951). The probe noise amplitude under these conditions was found in the present experiments to be dependent on the value of the impedance of the probe circuit. When this was sufficiently large for the lowest frequency present in both the probe and the tube circuits the appearance and amplitude of noise in both circuits were identical. Using large impedance circuits, comparisons of appearance and amplitude were made along the non-saturated and saturated sections of the discharge characteristic curve for various values of filament current. Different positions along and across the discharge were tried and different areas and orientation of the probe plates were used. In all cases the oscilloscope records for noise across the tube and for a probe at space-potential agreed closely in both appearance and amplitude.

The fact that the noise is a maximum for a probe at space potential may be used to find the latter rapidly. It was established in this way, and checked by conventional probe measurements, that the anode in this tube was invariably positive to the plasma, when the discharge was not current saturated. It was shown in similar ways that potential differences between different points in the plasma were always less than a volt.

Although the results show that the noise observed below saturation is not transmitted by plasma ion waves, they are not at first sight incompatible with transmission by the primary electrons. This is considered unlikely, however, because the noise amplitude was independent of probe position and orientation, and was unaffected by interposing a mica sheet between the cathode and a probe. It is concluded that low frequency noise in a non-saturated discharge is a coherent, random fluctuation of the whole plasma. This fluctuation will be equally effective across the discharge electrodes and at a probe anywhere in the plasma at space potential.

To test this idea, a controlled fluctuation of the plasma space potential was created and its effect studied. Space potential fluctuations of about 1 v were produced with a repetition rate of the order of a few hundred or thousand cycles per second by using one of the probes as a control electrode and applying a sinusoidal potential of this frequency to it. The effect of this on the current to another probe was studied for different values of the potential of the latter. At space potential the second probe showed a maximum fluctuation amplitude; this agreed in appearance and amplitude with that appearing across the tube. At other potentials the presence of an electron or ion sheath attenuated the probe current fluctuations, an ion sheath proving again more effective than an electron one.

A ' floating ' probe, or one connected to the anode only by a measuring instrument of high impedance, will assume almost the potential of the tube wall. This was about 20 v negative to space potential so that a sheath of ions will form at the probe, confining the field close to it. When this sheath was disturbed it was found that a new type of noise appeared. Its wave-form was coarser than that previously referred to, and it was confined to the anode–probe circuit and not transmitted to the anode–cathode circuit nor to other probe–anode circuits. The new probe noise could be as large as 1 v in amplitude. Two methods of disturbing the space-charge sheath were used : (*a*) bombardment with 50 to 100 v electrons from the cathode under conditions of tube current saturation, and (*b*) making use of the natural plasma potential fluctuation of the tube under non-saturated conditions. The results obtained by both methods were the same.

§ 4. SUMMARY AND DISCUSSION

This work has confirmed and extended the experimental results of previous workers, although some of the results obtained remain incompletely explained. Anode glow disturbances have been eliminated and low frequency noise isolated for study. The plasma space potential in the tube used was found to be negative to the anode when the discharge was not current saturated.

In a non-saturated discharge, the instantaneous occurrence of tube noise throughout the plasma, its high propagation speed, the good agreement between probe and tube noise and the absence of any directional effects lead to the conclusion that the plasma space potential was fluctuating coherently. This was confirmed by experiments under controlled conditions.

Essentially, the plasma of the discharges studied can be considered as a virtual anode with the tube voltage drop taking place across the cathode double space-charge sheath. Should this sheath fluctuate, the potential difference across it will vary, thus giving rise to a similar variation of the plasma space potential. The noise in the discharges, therefore, could be closely related to the stability of the cathode sheath.

Under such conditions an internal probe is most affected when it is at space potential. At other probe potentials the existence of a space-charge sheath at the probe produces attenuation and a phase change.

In the case of a ' floating ' probe, equilibrium of its space-charge sheath is disturbed somewhat differently by this space potential fluctuation and a coarser noise recorded. The same effect is obtained at a ' floating ' probe if the equilibrium of its space-charge sheath is disturbed by primary electron bombardment. This coarser noise may therefore be due partly to vibration or fluctuation of the ion sheath at the probe itself. Since the wall of a tube is covered on the inside by an ion sheath, the latter will govern to a considerable extent the response of an external probe to noise fluctuations in the plasma.

ACKNOWLEDGMENTS

This work was done under the supervision of Mr. T. P. Allen and Professor K. G. Emeléus whilst the author was a Queen's Scholar of the University of Malaya. The author is grateful to the members of staff of the Physics Department and the Radio Engineering Department of Queen's University, Belfast, for their help and co-operation.

References

Appleton, E. V., and Childs, E. C., 1930, *Phil. Mag.*, **10,** 969.

Armstrong, E. B., Emeléus, K. G., and Neill, T. R., 1951, *Proc. R. Irish Acad.* A, **54,** 291.

Cobine, J. D., and Gallagher, C. J., 1947, *J. Appl. Phys.*, **18,** 110.

Martin, H., and Woods, H. A., 1952, *Proc. Phys. Soc.* B, **55,** 281.

Terman, F. E., Buss, R. R., Hewlett, W. R., and Cahill, F. C., 1939, *Proc. Inst. Radio Engrs, N.Y.*, **27,** 649.

Tonks, L., and Langmuir, I., 1929, *Phys. Rev.*, **33,** 195.

Hysteresis of the Angle of Contact of Mercury against Steel

By D. J. WRIGHT*
University of Nottingham

Communicated by L. F. Bates ;
MS. received 5th *November* 1954 *and in amended form* 3rd *January* 1955

Abstract. Experiments have been made to investigate some of the causes of the hysteresis of the angle of contact of mercury against steel. It is shown by experiments *in vacuo* and in gases at low pressures, that adsorption on the solid surface is a principal cause of hysteresis. In contrast to the results of earlier work, however, roughness does not affect the hysteresis, though the angles of contact themselves are altered. Some agreement with theoretical predictions is obtained.

§ 1. Introduction

The first observation of the fact that the angle of contact between a liquid and a solid depends on the direction of motion of the liquid relative to the solid surface was made by Rayleigh (1890), and subsequently confirmed by Pockels (1914). Sulman (1920) termed the phenomenon 'hysteresis' and since then several theories have been proposed to account for it.

§ 2. Theoretical Discussion

Since the angle of contact appears to depend upon whether the liquid moves over a previously wetted surface or not, it is clear that the conditions of the solid surface must be a contributory factor. In the first place the roughness of the surface will have a definite effect on the angle of contact. Bartell and Smith (1929) and Wenzel (1936) considered this aspect, and the latter proposed the formula $\cos\theta' = r\cos\theta$ where θ' is the observed angle of contact, θ the theoretical angle of contact with a smooth solid, and r the roughness factor, defined as the ratio of the actual surface area to the geometric surface area.

Irregularities on the surface of the solid appear to play the role of small capillaries and their effect has been shown by Lee (1936) with tar on rough stones, by Fogg (1947) with water on leaves, and by others. Doss and Rao (1938) showed that a theoretical interpretation similar to the empirical relation of Wenzel could be obtained. Shuttleworth and Bailey (1948) concluded from a theoretical investigation that "the inevitable roughness of solids is sufficient to explain hysteresis", and that in fact there existed a unique equilibrium angle of contact. Bicerman (1950) endeavoured to make a practical determination of the variation of the angle of contact of water on steel plates of varying roughness as determined by means of a tracer instrument. Results of his early experiments appear inconclusive but serve to show the difficulties in obtaining quantitative confirmation of theory. This emphasizes the view expressed by Cassie (1948) that the variation in published values of θ is due to the difficulty in obtaining comparable surfaces.

---

* Now at University College, Auckland, New Zealand.

Adsorption on to the solid surface is an undoubted cause of hysteresis. However, here the problem is complicated by the type of adsorption and the phase state of the adsorbed molecules. Bangham and Razouk (1937) consider that the adsorbed molecules differ from the bulk liquid. They picture the adsorbed film of liquid vapour on the solid as a Langmuir type monolayer, the molecules of the film having fixed sites. Cassie assumes that molecules are adsorbed in clusters on single sites and that the film is altered by passage of the bulk liquid over it. In this way a qualitative explanation of the fact that the 'advancing' angle is greater than the 'receding' angle is given. This theory predicts that the receding angle may have various values while the advancing angle is unique. This conclusion is in contradiction to the theory of Yarnold and Mason (1949) who extended the methods of Doss and Rao in an attempt to allow for the adsorption of air as well as of vapour and also for an effect due to the time of immersion of the solid. Their conclusions indicate that the possible ranges of values for the two angles are similar.

Attempts to obtain decisive experimental evidence on this point have until now been inconclusive. In this work the problems undertaken were to investigate how the hysteresis of the angle of contact varied with adsorption from the atmosphere in which the solid was placed, and to attempt some qualitative experiments on the effects of roughness. To this end the hysteresis of the angle of contact of mercury against various specimens of steel was measured *in vacuo*, in hydrogen and in carbon dioxide at low pressures and in air at atmospheric pressure.

§ 3. EXPERIMENTAL DETAILS

Several methods of measuring contact angles have been devised. It is, however, apparent that it is not always the same quantity that is being measured. Claims have been advanced for the existence of at least seven different 'angles of contact' classified under dynamic or static, advancing or receding, and equilibrium angles. Only by the method of Yarnold (1946) does it seem possible to measure a true advancing or receding angle of contact and thereby investigate the hysteresis.

Yarnold's method consists in determining the upthrust on a spherical specimen as it is gradually immersed in a liquid (figure 1). Let U be the total upward force on a sphere of radius a immersed in a liquid so that its centre is at height y above the plane of the liquid surface. Let U_1 be the corresponding force for complete immersion.

It may be shown that

$$\frac{U}{U_1}=\frac{1}{2}-\frac{3y}{4a}+\frac{1}{4}\left(\frac{y}{a}\right)^3 \qquad \ldots\ldots(1)$$

when the liquid surface is accurately horizontal to the surface of the sphere, and $-\cos\theta=y/a$. This condition may be determined by the intersection of a theoretical curve obeying equation (1) with the experimental curve of U/U_1 against y/a.

As in Yarnold's original method a Sucksmith ring balance was used also in the present experiments to measure the upthrust U. The ring balance was made of spring steel, of cross section 0·58 cm by 0·01 cm, bent into a circle of radius 5·2 cm. Observations were made by means of telescope and scale using two mirrors attached to the ring in the usual positions.

To measure *y*, the depth of immersion of the sphere was obtained by observing the movement of the profile of the mercury surface over the eyepiece scale of a travelling microscope. A correction had to be applied for the vertical movement of the lowest point of the ring. This correction was found to be linear.

An improvement to the original method of altering the level of the liquid surface consisted in raising a reservoir of the liquid on a piston operated by a phonic motor, controlled by a beat frequency oscillator. This ensured a continuous even movement of the liquid surface and allowed variations to be made in the rate of advance. Preliminary experiments indicated that the balance itself was free from any hysteresis effects. For example, at the instant at which the sphere became free from the mercury surface, the reading of the balance returned to its zero without any drift. In order to obtain a true value of the position of the liquid surface when first contact with the sphere was made, it was necessary to take readings through the microscope and telescope at regular intervals of time, whence by interpolation the starting position could be accurately defined in relation to the instant of time when contact was first made.

The ring balance was contained in a brass box fitted with a lid (figure 1) and connected to the remainder of the apparatus by a ground joint. Vibration of the balance was minimized by a small damping vane which moved in a dash pot filled with Apiezon oil 'B' of low vapour pressure.

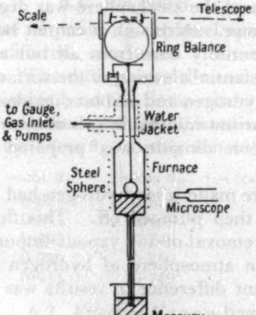


Figure 1. Apparatus for the determination of the angle of contact between a steel ball and a moving mercury surface *in vacuo* (Yarnold's method).

The main portion of the glass apparatus consisted of a wide tube extending below the balance box, and connecting through a capillary tube to the mercury reservoir. The wide tube was of 5 cm diameter, thus ensuring that the liquid surface at the centre could be considered plane, following the work of Gibson (1932). A side tube made connection to a McLeod gauge, the gas inlet system and the oil rotary pumps. A furnace wound in two half cylinders was fitted so that it could be swung into position around the wide tube for degassing purposes. A water jacket kept the remainder of the apparatus cool when the furnace was in use. The length of the capillary tube was such that when the apparatus was evacuated the level of the mercury surface was still within the capillary. As the reservoir was raised a fresh surface could then be formed by overflowing. According to Burdon (1935), a mercury surface formed at a pressure of 10^{-4} mm does not take up an adsorbed layer.

§ 4. PREPARATION OF SPECIMENS

The steel specimens used for these experiments were $\frac{1}{2}$ in. ball bearings. The treatment of each specimen was similar, so that any alteration of surface characteristics is of little importance for comparative purposes. Before use, the ball bearings were washed in carbon tetrachloride, several times in methyl alcohol, and finally in ether of AR standard. For some of the experiments a high degree of polish was obtained with jewellers rouge, and electrolytic polishing was also tried. Spheres used were carefully measured after polishing to ensure that there was no departure from a spherical shape. The mercury used was of AR standard. When required, purification was carried out by a standard method of washing with dilute nitric acid, and repeated distillations under vacuum. All the glass portions of the apparatus were cleaned by washing with a concentrated chromic acid solution followed by distilled water and thorough drying.

With the steel sphere in position, hanging in the centre of the wide tube, the apparatus was evacuated. The furnace was placed in position and the main portion of the apparatus and the steel sphere was degassed by heating to 500°c for three quarters of an hour. Although it cannot be claimed that this process leaves the solid surface entirely bare from all but adsorbed mercury vapour, it is considered that a substantial cleaning of the surface is achieved.

The gases used were hydrogen and carbon dioxide. Hydrogen was prepared by the electrolysis of a barium hydroxide solution, dried and stored over phosphorus pentoxide. Carbon dioxide was prepared by heating AR sodium bicarbonate.

A few experiments were made when hydrogen had been introduced following the initial degassing and then pumped off. This flushing out procedure was attempted to help in the removal of any vapour impurities so that the specimen had effectively been in an atmosphere of hydrogen immediately prior to the experiment. No significant difference in results was found by this means and the procedure was abandoned.

§ 5. ACCURACY

The accuracy of the method depends both upon the accuracy of readings taken with the telescope and microscope and upon noting the instant at which contact is made between the solid and the liquid. Since the rate at which the cosines of the angles vary depends upon the angle, the advancing angle was accurate to $\pm\frac{1}{4}°$, and the receding angle to $\pm\frac{1}{2}°$ in any one experimental run. Since there was no lag in the balance, the values of the angles of contact obtained were those of the true advancing and receding angles. The experimental curve (figure 2) commences below the y/a axis due to the initial adhesion of the mercury to the sphere. Similarly, for the receding angle, the experimental curve continues below the same axis, though at a different value, as the characteristic neck due to adhesion forms when the surface tension forces act downward. The bead of mercury reported by Yarnold as remaining attached to the sphere after contact between the liquid and solid had been broken was not observed in this work. The inference is that in this case the solid surface was at least freer of impurities than in the former work.

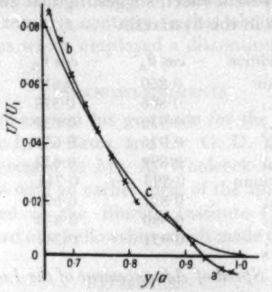


Figure 2. Typical experimental curves : a, liquid advancing ; b, liquid receding. Theoretical curve : c, representing the relation $U/U_1 = \frac{1}{2} - \frac{3}{4} y/a + \frac{1}{4}(y/a)^3$. The points of intersection give the angles of contact from the relation $-\cos\theta = y/a$.

§ 6. Results and Discussion

6.1. *Roughness*

Due to the varying state of polish of the steel specimens it was difficult to compare results between different specimens. Some qualitative comparisons are enlightening, however. The table shows the cosines of the angles obtained with two unpolished specimens, and for comparison the values obtained with three polished specimens. Both angles are affected individually by roughness, but the hysteresis, as given by $-(\cos\theta_A - \cos\theta_R$ is not significantly affected.

Specimen	$-\cos\theta_A$	$-\cos\theta_R$	$-(\cos\theta_A - \cos\theta_R)$ mean
1. Polished	0·884	0·624	0·199
	0·891	0·699	
	0·893	0·748	
2. Unpolished	0·980	0·810	0·186
3. Unpolished	0·951	0·749	

6.2. *Effect of Vacuum*

With the apparatus slightly modified to allow experiments to be made at atmospheric pressure in air, results could be compared with those obtained with the same specimen *in vacuo*. The general result obtained was that the difference in the cosines fell by about one third *in vacuo*. This indicates that a large proportion of the normally observed hysteresis is in fact due to adsorption of a layer of gaseous molecules on to the solid. It cannot be claimed that this layer was completely removed in the present experiments *in vacuo* but a substantial clean-up of the surface was obtained. The reduction in the hysteresis showed itself mainly as a large change in the value of $\cos\theta_R$.

6.3. *Hysteresis in Gases*

Compared to vacuum conditions, the hysteresis of the angle of contact increased when gas was admitted to the system. Pressures used ranged from 0·005 mm to about 1 mm and no significant variation with pressure could be found.

The magnitude of the hysteresis was greater with carbon dioxide than with hydrogen, suggesting that the carbon dioxide forms an adsorbed layer on the solid more easily. Forming the mercury surface either before or after the gas was admitted had no apparent effect, suggesting that adsorption on to the liquid surface played little part in the hysteresis.

Observer	Condition	$-\cos\theta_A$	$-\cos\theta_R$	$-(\cos\theta_A - \cos\theta_R)$
Yarnold	Air	0·860	0·610	0·250
,,	,,	0·806	0·575	0·231
,,	,,	0·930	0·624	0·306
Wright	,,	0·916	0·655	0·261
,,	,,	0·899	0·629	0·270
,,	Vacuum	0·893	0·748	0·145
,,	,,	0·891	0·699	0·192
,,	,,	0·980	0·810	0·170
,,	,,	0·951	0·749	0·202

6.4. *Speed of Advancement of the Liquid*

No differences were detected in any of these experiments when the speed of advancement or recession of the liquid surface was altered by a factor of two by means of the beat frequency oscillator, which controlled the motion of the mercury reservoir. The speeds used were of the order of two millimetres per minute.

Specimen	Conditions	$-\cos\theta_A$	$-\cos\theta_R$	$-(\cos\theta_A - \cos\theta_R)$
1.	Vacuum	0·893	0·748	0·145
	Hydrogen	0·886	0·699	0·187
	Carbon dioxide	0·891	0·646	0·245
2.	Vacuum	0·883	0·699	0·184
	Hydrogen	0·900	0·680	0·220
3.	Vacuum	0·917	0·688	0·229
	Carbon dioxide	0·940	0·681	0·259

6.5. *Consistency of Results*

A satisfactory aspect of the work was the reproducibility of the results obtained with the same specimen under similar conditions. For example, the variation in $-\cos\theta_A$ for one specimen was only 0·001 over a series of several determinations.

The values of the two quantities $-\cos\theta_A$ and $-\cos\theta_R$ measured for the same specimen under different conditions was also of interest: $\cos\theta_A$ proved to be constant to $\pm 1\%$ over a series of sixteen determinations, while $\cos\theta_R$ varied over a much greater range. This is of interest in conjunction with Cassie's theory, that $\cos\theta_A$ has a unique value.

The results quoted clearly indicate the difficulties to be met when attempting to compare the results from different specimens directly. However, that does not render invalid the quantitative conclusions that may be drawn from the study of the grouped results.

§ 7. CONCLUSIONS

(i) The hysteresis of the angle of contact of a liquid with a solid has been shown to be due in part to the adsorption on to the solid surface of gases from the atmosphere in which the solid is placed. This hysteresis may be reduced by degassing and measuring the angles *in vacuo*.

(ii) In agreement with the theory of Wenzel, it has been found that roughness of the solid affects the absolute values of the angles measured. However, contrary to the view expressed by Shuttleworth and Bailey, the variation of roughness used in these experiments did not affect the magnitude of the hysteresis.

(iii) The consistency in the values of the advancing angle under different conditions lends credence to the theory of Cassie that the advancing angle has a unique value. This result is contrary to that obtained by Yarnold with an earlier form of apparatus which employed a discontinuous alteration of level of the liquid surface.

ACKNOWLEDGMENTS

The author wishes to express his gratitude for the help and encouragement received from Professor L. F. Bates, and Dr. G. D. Yarnold who initiated this work. Thanks are expressed to Mr. A. Woolcock for his assistance and the benefit of his experience with an earlier form of the apparatus used in this work. The author is indebted to the Shirley Institute (British Cotton Research Association) for the award of a Fellowship which made this investigation possible.

REFERENCES

BANGHAM, D. H., and RAZOUK, R. I., 1937, *Trans. Faraday Soc.*, **33**, 1459.
BARTELL, F. E., and SMITH, C. N., 1929, *Industr. Engng Chem.* (*Industr.*), **21**, 1102.
BICERMAN, J. J., 1950, *J. Phys. Chem.*, **54**, 653.
BURDON, R. S., 1935, *Proc. Phys. Soc.*, **47**, 460.
CASSIE, A. B. D., 1948, *Disc. Faraday Soc.*, **3**, 11.
DOSS, K. S., and RAO, B. S., 1938, *Proc. Indian Acad. Sci.*, **A7**, 113.
FOGG, G. E., 1947. *Proc. Roy. Soc.* B, **134**, 503.
GIBSON, E. F. H., 1932, *Trans. Roy. Soc. S. Aust.*, **56**, 51.
LEE, A. R., 1936, *J. Soc. Chem. Ind., Lond.*, **55**, 23T.
POCKELS, A., 1914, *Phys. Z.*, **15**, 39.
RAYLEIGH, Lord, 1890, *Phil. Mag.*, **30**, 397.
SHUTTLEWORTH, R., and BAILEY, G. L. J., 1948, *Disc. Faraday Soc.*, **3**, 16.
SULMAN, H. L., 1920, *Trans. Instn. Min.* (*Metall.*) *Engrs*, **29**, 88.
YARNOLD, G. D., 1946, *Proc. Phys. Soc.*, **58**, 120.
YARNOLD, G. D., and MASON, B. J., 1949, *Proc. Phys. Soc.* B, **62**, 121.
WENZEL, R. N., 1936, *Industr. Engng Chem.* (*Industr.*), **28**, 988.

The Heat Changes Accompanying Magnetization Processes in Ferrites

By L. F. BATES and N. P. R. SHERRY
University of Nottingham

MS. received 13*th December* 1954

Abstract. Measurements have been made, for the first time, of the thermal changes accompanying the step-by-step magnetization of a number of ferrites. The method was that used with metals by Bates and his co-workers, with the added improvement that the thermal deflections were recorded via a split photocell amplifier.

Measurements were made on certain mixed ferrites of the nickel–zinc series. The results showed that the major part of the heat changes can be explained satisfactorily in terms of the changes of intrinsic magnetization with field. In some of the specimens a slight irreversible heating was observed ; it therefore appears that domain rotational effects make a very small contribution to the thermal changes. An anomalous, low field effect previously found in certain metals occurred in some of the ferrites. This effect is not yet fully understood, although some qualitative suggestions are made to account for it ; it is concluded that the observed thermal effects in ferrites can be explained on the same lines as those for metals.

§ 1. Introduction

When the magnetic field applied to a ferromagnetic material is changed there are, in addition to the change of magnetization, changes in the magnetic and thermal energies of the material. These result in a small change of the temperature of the specimen (the magnetothermal effect). Much has been done in recent years to make accurate measurements of these temperature changes as the ferromagnetic materials are taken step-by-step around a hysteresis cycle (Bates 1951). Such investigations may be of value in the analysis of magnetization curves, particularly from the point of view of associating the observed energy changes with particular magnetization processes.

§ 2. Thermal Measurements on Ferrites

The rapid development and wide application of ferrites make it important to perform magnetothermal investigations on them. A preliminary experiment on a nickel–zinc ferrite (Bates and Sherry 1953) showed that the measured heat changes varied linearly with the change in magnetizing field over almost the whole range of a hysteresis cycle. These measurements suggested that the only outstanding energy process which occurred was that associated with the usual, small, change of intrinsic magnetization with magnetizing field. This effect can be calculated from the simple thermodynamic relation

$$\Delta Q' = -T\frac{\partial I_0}{\partial T}\Delta H \qquad \ldots\ldots(1)$$

in which $\Delta Q'$ is the energy change corresponding to a field change ΔH, I_0 is the intrinsic magnetization and T the absolute temperature. Numerical calculations employing this formula showed reasonable agreement with the measured heat changes.

Now, it would appear from this single investigation that magnetothermal measurements were of only limited interest in the case of ferrites. It was realized, however, that the low Curie point (104°c) of the particular 15% NiO, 35% ZnO, 50% Fe_2O_3 ferrite examined would result in such a large value of $\partial I_0/\partial T$ that the energy term corresponding to equation (1) would probably be sufficiently great to obscure any other processes that might occur. Accordingly, measurements were made on a series of nickel–zinc ferrites of different compositions. In this series of materials (ferroxcube B) an increased percentage of NiO is accompanied by a rise in the Curie temperature. The specimens used were commercial samples in the form of rods some 15 cm long and 5 mm in diameter.

The heat changes accompanying magnetization of the ferrites were measured using the apparatus previously employed for similar measurements on silicon–iron (Bates and Marshall 1953), but a noteworthy improvement was made in the method of observing the deflections of the sensitive long period galvanometer. Previously, optical magnification had been obtained by using a telescope with a scale set at a distance of eight metres from the galvanometer. This system was now replaced by a split photocell amplifier similar to that described by Preston (1946), with the exception that negative feedback was not employed. The use of the amplifier resulted in a slight improvement of the overall sensitivity of the apparatus but gave the great advantage that the thermal deflections could be observed much more easily.

The nickel–zinc ferrites had the following approximate compositions:

Specimen 1.	15% NiO	35% ZnO	50% Fe_2O_3
2.	25% NiO	25% ZnO	50% Fe_2O_3
3.	39% NiO	11% ZnO	50% Fe_2O_3
4.	48% NiO	2% ZnO	50% Fe_2O_3

The results for specimen 1 were reproduced in the graph already published (Bates and Sherry 1953). The result for the remaining specimens are shown graphically in figures 1 to 3. In these graphs the total energy changes Q' are

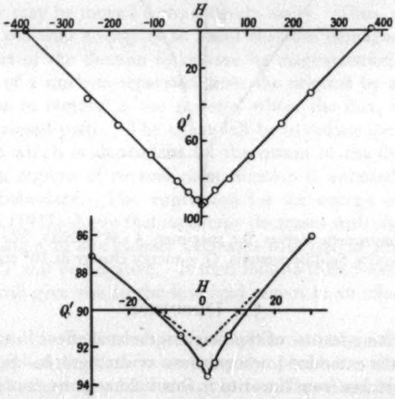


Figure 1. Thermomagnetic curves for specimen 2 : 25% NiO, 25% ZnO, 50% Fe_2O_3. H=effective field in oersteds, Q'=Energy change in 10^3 erg cm^{-3}.

plotted as usual against the effective field H. The value of the thermal sensitivity adopted in these measurements was obtained by suddenly applying a known tensile force to the specimen.

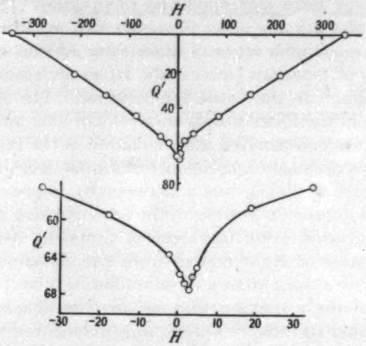


Figure 2. Thermomagnetic curves for specimen 3 : 39% NiO, 11% ZnO, 50% Fe_2O_3. H=effective field in oersteds, Q'=energy change in 10^3 erg cm^{-3}.

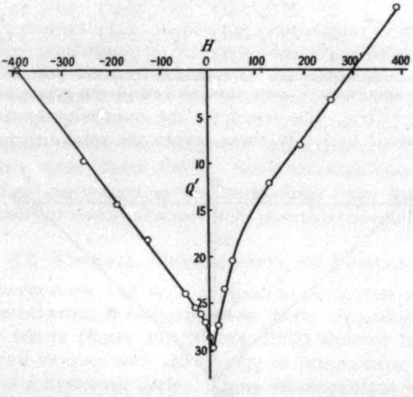


Figure 3. Thermomagnetic curves for specimen 4 : 48% NiO, 2% ZnO, 50% Fe_2O_3. H=effective field in oersteds, Q'=energy change in 10^3 erg cm^{-3}.

§ 3. DISCUSSION

The most striking feature of the magnetothermal effect in all the four ferrites investigated are the extended linear portions of each graph. Some of the graphs show slight departures from linearity in low field regions, but it is clear that the major cause of the liberation or absorption of heat is associated with change in intrinsic magnetization. This can be readily checked by applying the simple relation

$$Q' = -\frac{T}{I_0}\frac{dI_0}{dT}\int d(IH) \qquad \ldots\ldots(2)$$

first given by Stoner and Rhodes (1949). For this purpose measurements, on small samples of the materials, were made of the relative variation of intrinsic magnetization with temperature, using the well-known method of Weiss and Forrer. Reasonable agreement was found between the directly measured values and those calculated from the overall heat change, the calculated values in each of the four cases being the smaller.

In specimen 4 alone is there any marked irreversible process. In the above curves this appears in the form of a step, cf. figure 3, which is superimposed on the magnetocaloric effect in the low field region. The work of Rhodes (1948) suggests that this irreversible effect should be centred about the coercive field value. This is supported by the experimental results. As would be expected, the magnitude of the irreversible step is equal to the overall heating per half-cycle. Thus, the remainder of the cycle may be assumed to show only reversible effects.

The observations on ferrites have also given further examples of the interesting anomalous low field magnetothermal effect, ‘ the dip ’, previously noticed in work on iron and silicon–iron. If the linear portions of the thermal curves of the ferrites (specimens 2, 3 and 4) are extended to the low field region, it is seen that greater cooling and heating effects are present than would be expected on the grounds of the magnetocaloric effect alone. Careful investigation has shown that this phenomenon is real and not due to experimental error. At the moment no completely satisfactory explanation of ‘ the dip ’ in ferrites can be offered.

A qualitative explanation suggested by Bates and Marshall (1953) is based on a flux-closure mechanism postulated by Bozorth (1947), who pointed out that in a material which possesses only small strain and crystal anisotropy a domain boundary may be moved comparatively easily. Thus, when the applied field is removed the wall energy is so small that the demagnetizing field may tend to cause part of the domain to reverse its magnetization. This can take place by growth of a nucleus separated from the original by a 180° boundary, and may give rise to regions in the material where the flux, to a considerable extent, follows a closed path. The effect will be to reduce the overall magnetization to a value which is dependent on the extent of the flux closure. The tendency to form regions of reverse magnetization is opposed by the stability of the domain boundary. The expression for the energy of a domain wall as given by Bloch (1932) shows that its energy decreases with rise in temperature. Thus, the obstruction to flux closure decreases with rise in temperature so that the value of dI/dT will be negative. It then follows from thermodynamic arguments that this will give rise in the low field region to an effect which forms a dip.

Now, although the original argument applies strictly only in the absence of an external applied field, there remains the possibility that the effect may occur to a limited extent in small fields. In fact, Bates and Martin (1953) have shown by powder pattern experiments that the formation of demagnetization nuclei at imperfections takes place over a comparatively wide range of applied field. The appearance of these regions of reverse magnetization must give rise to an effective flux closure. In materials with low anisotropy the growth of reverse magnetization regions may be favoured by a rise in temperature. This implies, in just the same way as in the Bozorth explanation, that the sign of dI/dT will be negative and a dip may occur.

As a qualitative test of this explanation thermal readings were made on a material with low anisotropy. It is known that the nickel–iron series of alloys shows a minimum anisotropy in the region of 79% Ni, and therefore a permalloy rod was selected having a composition 78% Ni, 4% Mo and 18% Fe. The thermal results showed a large magnetocaloric effect with a clear 'dip' in the zero field region (figure 4).

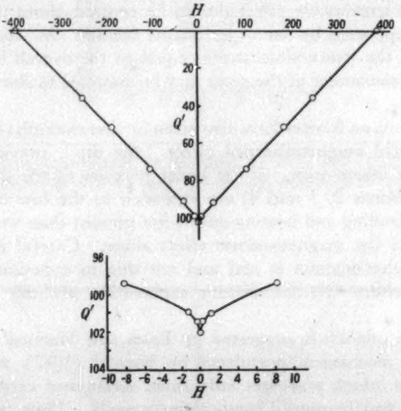


Figure 4. Thermomagnetic curves for a Permalloy specimen : 78% Ni, 4% Mo, 18% Fe. H=effective field in oersteds, Q'=energy change in 10^3 erg cm^{-3}.

The anomalous effect has been discussed extensively here as it is thought that the 'dip' may arise from some important magnetic process that is not yet fully appreciated or understood. For example, all the specimens used were polycrystalline, and free poles must have been present at the grain boundaries, and may have played a big part in the thermal changes in materials of such low coercivity and high initial permeability. Some experiments are now in hand in this laboratory to try to measure their effects and test the suggestions made in the above discussion.

The thermal contribution of rotational processes has been found to be very important in some metals. However, in the ferrites investigated here the agreement between the measured and calculated magnetocaloric effects is sufficiently good to show that the effect of rotational processes is very small, and it is therefore not possible to derive information about the value of the Stoner and Rhodes b'' coefficient in ferrites.

§ 4. Conclusion

It is seen from the new magnetothermal measurements on ferrites that, in spite of the great differences between the ferrites and normal ferromagnetic metals, the heat changes that occur are explicable in terms of the same basic causes, although the relative importance of the various contributory processes may be very different.

Acknowledgments

We wish to thank the General Electric Company, Schenectady, the Marconi Wireless Telegraph Company Limited and Philips Research Laboratories, Eindhoven, who provided rod specimens of ferrites suitable for the experiments.

Thanks are also due to Dr. A. W. Simpson for help in the measurements on the variation of intrinsic magnetization with temperature. One of us (N. P. R. S.) is indebted to the Department of Scientific and Industrial Research for a maintenance grant during part of the work.

References

Bates, L. F., 1951, *J. Phys. Radium*, **12**, 459.
Bates, L. F., and Marshall, G., 1953, *Rev. Mod. Phys.*, **25**, 17.
Bates, L. F., and Martin, D. H., 1953, *Proc. Phys. Soc.* A, **66**, 162.
Bates, L. F., and Sherry, N. P. R., 1953, *Proc. Phys. Soc.* B, **66**, 609.
Bloch, F., 1932, *Z. Phys.*, **74**, 329.
Bozorth, R. M., 1947, *Z. Phys.*, **124**, 519.
Preston, J. S., 1946, *J. Sci. Instrum.*, **23**, 173.
Rhodes, P., 1948, *Proc. Leeds Phil. Lit. Soc.*, **5**, 116.
Stoner, E. C., and Rhodes, P., 1949, *Phil. Mag.*, **40**, 481.

Carrier Concentration Disturbances in Semiconductors

By G. G. E. LOW
Department of Physics, University of Reading

Communicated by R. W. Ditchburn ; MS. received 13*th November* 1954
and in revised form 21*st December* 1954

Abstract. Carrier concentration disturbances in semiconducting filaments are discussed by reference to the composition of the current at the end contacts. Four types of disturbance must be envisaged, of which carrier injection is the best known. Carrier exclusion and extraction have also been experimentally observed, but carrier accumulation has not yet been reported. These phenomena are discussed in terms of a unified set of equations on the assumption that the disturbances are much smaller than the equilibrium concentration of majority carriers.

---

§ 1. Introduction

In the case of those semiconductors in which holes as well as electrons play a part in conduction it is possible to change the concentration of current carriers by adding or subtracting equal numbers of both types of carrier. Concentration disturbances of this kind which arise when current is passed through a contact placed on the surface of a semiconductor can be conveniently discussed by considering the conditions which prevail in an n-type semiconducting filament with suitable end contacts (figure 1 (*a*)). For simplicity, it will be assumed that the filament is long so that the carrier disturbances arising near the two contacts can be considered separately. Under equilibrium conditions, a fraction $\gamma_0=\mu_p p_0/(\mu_n n_0+\mu_p p_0)$ of the total current in the bulk of the semiconductor is carried by holes. The corresponding quantity which governs the composition of the current at the end contacts is the injection (or extraction) ratio, denoted by γ. If $\gamma=\gamma_0$ at both contacts, then the carrier concentration throughout the specimen remains undisturbed by current flow. In all other cases the concentrations in the neighbourhoods of the contacts are functions of the current. The condition $\gamma>\gamma_0$ at the contact biased positively with respect to the filament corresponds to the familiar case of *carrier injection*, whereas $\gamma<\gamma_0$ at the same contact must lead to a decrease of carrier concentration which will be called *carrier exclusion*. Correspondingly, at the contact biased negatively with respect to the filament, a decrease of carrier concentration is expected for $\gamma>\gamma_0$ and this is the phenomenon known as *carrier extraction*. The converse process, *carrier accumulation* is expected to occur at this same contact for $\gamma<\gamma_0$.

The purpose of the present note is to discuss these four types of disturbance in terms of a unified set of equations. The nomenclature of these phenomena is not yet generally standardized. The terms used in the following discussion are those defined above.

§ 2. Shockley's General Equations

Carrier recombination in a system such as that shown in figure 1 (*a*) has been discussed (Shockley 1950) for the case of concentration disturbances which are small compared with the equilibrium concentration of majority carriers. Under these conditions it can be assumed that the electric field in the filament is unaffected by the disturbances. Consequently, the components of the hole current density (i_p) can be written as

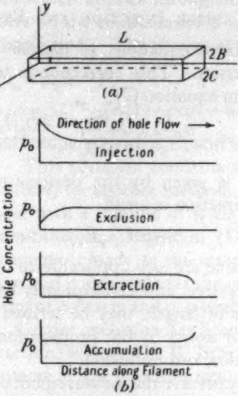


Figure 1. (*a*) Diagram showing the dimensions of the filament. The contact discussed lies in the plane $x=0$. (*b*) Plots showing the variation in hole concentration along the filament associated with each type of disturbance.

$$i_{px} = e\mu_p p E_x - eD_p \frac{\partial p}{\partial x} \qquad \ldots\ldots(1)$$

$$i_{py} = -eD_p \frac{\partial p}{\partial y} \qquad \ldots\ldots(2)$$

$$i_{pz} = -eD_p \frac{\partial p}{\partial z} \qquad \ldots\ldots(3)$$

where the symbols have their usual meanings. The continuity condition

$$\frac{\partial p}{\partial t} = \frac{p_0 - p}{\tau_b} - \frac{1}{e} \nabla i_p \qquad \ldots\ldots(4)$$

must be satisfied, where τ_b is the bulk lifetime and p_0 is the equilibrium concentration of holes. Substituting for i_p and putting $p_1 = p - p_0$ leads to

$$\frac{\partial p_1}{\partial t} = -\frac{p_1}{\tau_b} - \mu_p E_x \frac{\partial p_1}{\partial x} + D_p \nabla^2 p_1. \qquad \ldots\ldots(5)$$

A solution of this equation, symmetrical in y and also in z, is

$$p_1 = A \exp(-vt - ax) \cos by \cos cz \qquad \ldots\ldots(6)$$

where A is a constant and ν, a, b and c are constants related by

$$\nu + D_{\mathrm{p}}a^2 + \mu_{\mathrm{p}}E_x a = \frac{1}{\tau_b} + D_{\mathrm{p}}(b^2 + c^2) = \frac{1}{\tau_{\mathrm{f}}} \qquad \ldots\ldots(7)$$

(τ_{f} is defined as the filament lifetime).

§ 3. CARRIER INJECTION AND EXCLUSION

Shockley used the above equations to discuss the steady state condition arising from carrier injection. This corresponds to p_1 being independent of t, i.e. $\nu = 0$, and hence from equation (7)

$$a = \frac{-\mu_{\mathrm{p}}E_x \pm (\mu_{\mathrm{p}}^2 E_x^2 + 4D_{\mathrm{p}}/\tau_{\mathrm{f}})^{1/2}}{2D_{\mathrm{p}}}. \qquad \ldots\ldots(8)$$

The solution of interest is given by the positive sign. For $\mu_{\mathrm{p}}E_x \gg (D_{\mathrm{p}}/\tau_{\mathrm{f}})^{1/2}$ i.e. if diffusion in the x-direction is small

$$a \simeq 1/\mu_{\mathrm{p}}E_x\tau_{\mathrm{f}}. \qquad \ldots\ldots(9)$$

The equations derived above are also applicable to carrier exclusion, but in this case the constant A is negative, and consequently so is p_1. The quantity $1/a$, which has the dimensions of length, may be termed the down stream diffusion length as the electric field *assists* in the propagation along the filament of the disturbances due to injection and exclusion.

Experiments frequently involve the measurement of the change of conductance of the filament due to injection or exclusion. This can be found from the change in the total number of holes, which under steady state conditions is given by

$$P_1 = \int_0^L \int_{-B}^{B} \int_{-C}^{C} p_1\,dx\,dy\,dz = \int_0^L \int_{-B}^{B} \int_{-C}^{C} A \exp(-ax)\cos by \cos cz\,dx\,dy\,dz \qquad \ldots\ldots(10)$$

where ν has been put equal to zero. It is convenient to find an expression for A by considering equation (6) for $x = 0$ and, as before, $\nu = 0$. This leads to

$$A = \frac{p_1(0)}{\cos by \cos cz}. \qquad \ldots\ldots(11)$$

By differentiating equation (6) and substituting for $\partial p/\partial x$ in equation (1) the relation

$$i_{\mathrm{p}x} = \mathbf{e}\mu_{\mathrm{p}}(p_0 + p_1)E_x + \mathbf{e}D_{\mathrm{p}}ap_1 \qquad \ldots\ldots(12)$$

is obtained, and this may be used to find a value of $p_1(0)$ for substitution into equation (11). Thus combining equations (10), (11) and (12)

$$\begin{aligned} P_1 &= \int_0^L \int_{-B}^{B} \int_{-C}^{C} \frac{i_{\mathrm{p}x}(0) - \mathbf{e}\mu_{\mathrm{p}}p_0E_x}{\mathbf{e}(\mu_{\mathrm{p}}E_x + aD_{\mathrm{p}})} \exp(-ax)\,dx\,dy\,dz \\ &= \frac{I_{\mathrm{p}x}(0) - 4BC\mathbf{e}\mu_{\mathrm{p}}p_0E_x}{a\mathbf{e}(\mu_{\mathrm{p}}E_x + aD_{\mathrm{p}})}\{1 - \exp(-aL)\}, \end{aligned} \qquad \ldots\ldots(13)$$

where $I_{\mathrm{p}x}(0)$ is the total hole current at $x = 0$. This is given by γI where I is the total current through the filament. A simplification may be brought about by observing that

$$4BC\mathbf{e}\mu_{\mathrm{p}}p_0E_x = \gamma_0 I \qquad \ldots\ldots(14)$$

so that if

$$\gamma_{\mathrm{e}} = \gamma - \gamma_0 \qquad \ldots\ldots(15)$$

$$P_1 = \frac{\gamma_{\mathrm{e}} I}{a\mathbf{e}(\mu_{\mathrm{p}}E_x + aD_{\mathrm{p}})}\{1 - \exp(-aL)\}. \qquad \ldots\ldots(16)$$

The ratio γ_e may be called the effective injection ratio and it is this quantity which is actually measured experimentally by the methods described by Shockley, Pearson and Haynes (1949) and by Many (1954).

If diffusion in the x direction is negligible, the change of conductance of the filament can be found as follows. It has been shown (Shockley 1950) that, provided $p_1 \ll n_0$ (a condition which applies throughout the present discussion), the effect of additional carriers on the conductance is independent of the manner in which these carries are distributed. The change of conductance is then given by

$$\begin{aligned}\Delta G &= \mathbf{e}(\mu_n+\mu_p)L^{-2}P_1\\ &= (\mu_n+\mu_p)L^{-2}\gamma_e I\tau_f\{1-\exp(-aL)\} \qquad \ldots\ldots(17)\end{aligned}$$

where aD_p has been neglected compared with $\mu_p E_x$ and where a has been eliminated by the use of equation (9). This equation confirms a similar expression which has been derived by Many (1954) by another method, based on the assumptions that injected holes recombine with a lifetime of τ_f and that the filament is long enough to permit this recombination. Equation (17) does not involve these assumptions. For long filaments it leads to the same result as the steady state solution of Many's expression and it also includes the case of short filaments.

Carrier injection, both by point contacts and p–n junctions, has been extensively discussed in the literature. Carrier exclusion has been observed by Arthur *et al.* (1955) (called by them extraction) and by Dr. R. Bray (private communication), using specially prepared contacts of low injection ratio.

§ 4. CARRIER EXTRACTION AND ACCUMULATION

Provided that the sign of E_x is reversed the equations derived in § 2 apply equally to contacts biased negatively with respect to an n-type semiconductor. Under these conditions the steady state solution of equation (7) is

$$a' = \frac{\mu_p E_x \pm (\mu_p^2 E_x^2 + 4D_p/\tau_f)^{1/2}}{2D_p} \qquad \ldots\ldots(18)$$

where the positive sign corresponds to the solution of interest. It will be observed that the greater the relative importance of diffusion the greater is the distance to which the disturbance is propagated along the filament, and consequently the quantity $1/a'$ may be termed the up stream diffusion length as the electric field *opposes* the propagation. However, even if there is no electric field, the value of $1/a'$ is only $(\tau_f D_p)^{-1/2}$, i.e. a diffusion length. The total steady state change in the number of holes may be found by the same method as that used in the previous section. Consequently, provided that a is changed to a' and that the sign of E_x is reversed, equation (16) can also be applied to extraction and accumulation. On the other hand, as these phenomena are inappreciable when diffusion is negligible, the expression found for the change of conductance (equation (17)) is not applicable.

Carrier extraction was first envisaged by Bardeen and Brattain (1949). It has been observed by Banbury (1953) using evaporated contacts on germanium and lead sulphide, and by Angello and Ebert (1954) (called by them 'egression') using a p–n junction of indium alloyed to germanium. So far as is known, carrier accumulation has not been observed experimentally.†

---

† Since the preparation of this paper Dr. A. F. Gibson has informed the author that he has observed accumulation using an n–n junction in germanium.

Acknowledgments

The author wishes to thank Professor R. W. Ditchburn for his interest; Drs. H. K. Henisch and P. C. Banbury for encouragement and valuable discussions; and the New Zealand Defence Research Organisation for financial support.

References

Angello, S. J., and Ebert, T. E., 1954, *Phys. Rev.*, **96**, 221.

Arthur, J. B., Bardsley, W., Brown., M. A. C. S., and Gibson, A. F., 1955, *Proc. Phys. Soc.* B, **68**, 43.

Banbury, P. C., 1953, *Proc. Phys. Soc.* B, **66**, 50.

Bardeen, J., and Brattain, W. H., 1949, *Phys. Rev.* **75**, 1208.

Many, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.

Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

Shockley, W., Pearson, G. L., and Haynes, J. R., 1949, *Bell Syst. Tech. J.*, **28**, 335.

RESEARCH NOTES

Temperature Dependence of the Dielectric Constant of Diamond

By P. T. NARASIMHAN
Department of Physics, Indian Institute of Science, Bangalore 3

Communicated by R. Fürth ; MS. received 8th November 1954

Abstract. The temperature dependence of the dielectric constant of diamond has been measured over the temperature range 50–200°c. The value of $\epsilon^{-1}d\epsilon/dT$ over this range is $+1\times10^{-5}$. Details of the method of measuring the temperature coefficient of dielectric constant are also given. The magnitude and sign of $\epsilon^{-1}d\epsilon/dT$ for diamond has been theoretically calculated using Maxwell's relationship and Kramers–Heisenberg theory. The agreement between theoretical and experimental values is extremely good.

§ 1. Introduction

In the course of their extensive measurements on the dielectric constant of diamond, Whitehead and Hackett (1939) noticed that the value showed a slight increase with temperature at about 60°c. This increase was of the order of 1·5 to 2·5%, but with the limited accuracy of their apparatus they could only point out that the dielectric constant of diamond is substantially independent of temperature and thus in accordance with the Lorentz theory. The value of $5{\cdot}68\pm0{\cdot}03$ obtained by these authors is identical with the square of the refractive index for infinite wavelength, n_∞. Thus Maxwell's relationship $\epsilon=n_\infty{}^2$ (ϵ is the static dielectric constant) is very well obeyed (see also Fröhlich 1949). Now, accurate determinations of the variation of the refractive index of diamond at optical frequencies (Ramachandran 1947) show that the value of dn/dT is appreciable and has a positive value at temperatures above −100°c. From the order of magnitude of the dn/dT values it is evident that the variation of the dielectric constant of diamond with temperature will probably be of the order of only 10^{-5}. Even in the case of ionic crystals the values of $\epsilon^{-1}d\epsilon/dT$ are known to be of the order of only 10^{-4} (Bretscher 1934, Eucken and Büchner 1935). The slight increase in the value of the dielectric constant observed by Whitehead and Hackett was explained by Hartshorn (see the discussion following the paper of Whitehead and Hackett 1939) as due to the increased conductivity of the liquid mixture used in their measurements. In fact with the apparatus of Whitehead and Hackett the true effect could not have been measured at all owing to the very small order of magnitude. In the present work we have been able to measure the value of $\epsilon^{-1}d\epsilon/dT$ of diamond in the temperature range 50–200°c by means of a heterodyne beat apparatus and have found that the value is indeed of the order of 10^{-5} and is positive.

Ramachandran (1947) has derived a quantitative expression for the value of dn/dT on the basis of Kramers–Heisenberg theory from a knowledge of the characteristic frequencies and strength of the harmonic oscillators (ν_{0k} and f_{0k} respectively). Since Maxwell's relation has been known to hold good in the case of diamond it is possible to calculate the temperature coefficient of the dielectric constant of diamond on lines similar to Ramachandran's. Remarkable agreement between the theory and experiment has been found, as will be seen presently.

§ 2. EXPERIMENTAL DETAILS AND THE VALUE OF $\epsilon^{-1}\,d\epsilon/dT$

The changes in the dielectric constant were measured by changes in the capacity using the heterodyne beat apparatus which has been already described elsewhere (Narasimhan 1953). The voltage supply for the entire unit was taken from a constant voltage transformer. The frequency stability of the apparatus was remarkable (one beat per 10 seconds could be maintained for a period of several hours). For the measurement of the small capacity changes due to temperature an accurate variable air condenser of the stepped-rod type (Watson, Rao and Ramaswamy 1934) was used in parallel with the crystal condenser. This variable condenser had a total capacity of about 20 μF with a variable capacity of about 1·2 μF corresponding to about sixty revolutions of the graduated drum (360 divisions). A revolution counter attached to this drum made the readings easier. The condenser was calibrated as usual, before use. In the circuit was also a precision variable air condenser of the vane type which served the purpose of getting the heterodyne note within the coverage of the Watson condenser. After this adjustment, the precision condenser was kept ' locked '.

The diamond used in the present measurements was in the form of a nearly rectangular plate (length 6 mm, breadth 4·5 mm, thickness 1·1 mm, weight 69·4 mg). The two broader surfaces of this plate were aluminized fairly thickly so as to form the two plates of the crystal condenser. The crystal cell consisted of a brass container (3 in. in diameter, $\frac{5}{16}$ in. thick, $2\frac{1}{2}$ in. high) with a removable lid screwed on top. It was first silvered and then given a thick coating of gold in the interior. In the inside bottom of the cell was mounted a circular piece of ceramic insulator (2 in. in diameter, $\frac{1}{2}$ in. thick) on which was screwed a brass disc ($1\frac{1}{2}$ in. in diameter, $\frac{1}{8}$ in. thick) carrying a terminal for connecting the radio-frequency lead. This brass disc was also gold coated. The aluminized diamond plate was placed over this disc while a thick earthed metal probe contacted the other side of the diamond plate. The lead from the brass disc was taken out through a hole ($\frac{1}{2}$ in. in diameter) made at the side of the outer brass vessel. Covering this hole was a ceramic insulator carrying a terminal to which was connected the lead from the brass disc. The crystal cell was connected to the heterodyne apparatus by a shielded lead, the brass vessel forming the earthed electrode. The crystal cell could be heated to desired temperatures by means of a wire resistance heater strip placed at the outside bottom of the cell.

Temperatures were measured both by a mercury thermometer and a thermocouple at various points of the cell. The difference in temperature between the inner brass disc and the outer vessel was less than 0·5°C and thus it was found convenient to measure temperatures from outside. Regulation at desired temperatures could be effected not only by varying the input current to the heater strips but also by use of a bi-metallic strip arrangement. It was found desirable to work above room temperature since at lower temperatures the effect of moisture made the readings unreliable. Thus the temperature range 50–200°c was found convenient and the capacity changes were also within the coverage of the Watson condenser. The measurements involved in the first instance the accurate determination of the capacity change due to the crystal cell itself, i.e. without the diamond plate. This was obtained by a series of measurements over the required temperature range, the values being consistently reproducible. Measurements were then carried out with the diamond plate in the cell whereby the capacity C of the diamond plate and its variation with temperature dC/dT were obtained. The average of a series of determinations gave the value of

$$\frac{1}{C}\frac{dC}{dT} = +1{\cdot}1 \pm 0{\cdot}1 \times 10^{-5}$$

over the temperature range 50–200°c. According to Eucken and Büchner (1934)

$$\frac{1}{\epsilon}\frac{d\epsilon}{dT} = \frac{1}{C}\frac{dC}{dT} - \alpha$$

where α is the coefficient of linear expansion of the crystal. The thermal expansion of diamond has been accurately determined by Krishnan (1946 a) over a wide temperature range. From Krishnan's data the value of α is seen to be only of the order of 1 to 2×10^{-6} over the present range, and since the uncertainty in the value of $C^{-1}dC/dT$ is also of this order we may omit this correction here. In our measurements, the variation with temperature of $\epsilon^{-1}d\epsilon/dT$, if any, could not be measured, chiefly due to the small size of the diamond.

§ 3. DISCUSSION

If we assume with Whitehead and Hackett that the Lorentz relationship holds good in diamond, that is,

$$\frac{\epsilon - 1}{\epsilon + 2}\frac{1}{\rho} = \text{constant} \qquad \text{......(1)}$$

then the value of $\epsilon^{-1}d\epsilon/dT$ will be given by

$$\frac{1}{\epsilon}\frac{d\epsilon}{dT} = \frac{1}{\epsilon}\frac{\epsilon^2 + \epsilon - 2}{3}\frac{1}{\rho}\frac{\partial\rho}{\partial T}. \qquad \text{......(2)}$$

Now, since $\rho^{-1}\partial\rho/\partial T = -\gamma$ where γ is the coefficient of cubic expansion, we have

$$\frac{1}{\epsilon}\frac{d\epsilon}{dT} = -\frac{1}{\epsilon}\frac{(\epsilon^2 + \epsilon - 2)}{3}\gamma. \qquad \text{......(3)}$$

Thus we should expect a diminution in the dielectric constant with increase of temperature. The observed positive value cannot therefore be explained as a pure density effect. On the other hand we can calculate the value of $\epsilon^{-1}d\epsilon/dT$ on lines similar to those used by Ramachandran (1947) for the calculation of dn/dT. We have according to Kramers–Heisenberg theory

$$n^2 = 1 + \sum_k \frac{f_{0k}}{\nu_{0k}{}^2 - \nu^2}. \qquad \text{......(4)}$$

Since Maxwell's relation holds good in diamond we also have

$$\epsilon = n_\infty{}^2. \qquad \text{......(5)}$$

From this it follows that

$$\epsilon = 1 + \sum_k \frac{f_{0k}}{\nu_{0k}^2}. \qquad \ldots\ldots(6)$$

Differentiating (6) with respect to T we get

$$\frac{d\epsilon}{dT} = \sum_k \left\{ -\frac{2f_{0k}}{\nu_{0k}^3}\frac{d\nu_{0k}}{dT} + \frac{1}{\nu_{0k}^2}\frac{df_{0k}}{dT} \right\}. \qquad \ldots\ldots(7)$$

Following Ramachandran it can be shown that

$$\frac{1}{\nu_{0k}^2}\frac{df_{0k}}{dT} = -\gamma(\epsilon - 1).$$

Introducing the quantity

$$\chi_k = \frac{1}{\lambda_{0k}}\frac{d\lambda_{0k}}{dT} = -\frac{1}{\nu_{0k}}\frac{d\nu_{0k}}{dT}$$

we may write $\frac{d\nu_{0k}}{dT} = -\chi_k \nu_{0k}$.

Finally we get

$$\frac{1}{\epsilon}\frac{d\epsilon}{dT} = \frac{1}{\epsilon}\left\{ -\gamma(\epsilon - 1) + 2\sum_k \frac{f_{0k}\chi_k}{\nu_{0k}^3} \right\} \qquad \ldots\ldots(8)$$

or

$$\frac{1}{\epsilon}\frac{d\epsilon}{dT} = \frac{1}{\epsilon}\left\{ -\gamma(\epsilon - 1) + 2\sum_k f_{0k}\lambda_{0k}^2\chi_k \right\}. \qquad \ldots\ldots(9)$$

For diamond, Peter (1923) has given the following values: $k=1$, 2; $f_{01} = 3{\cdot}859 \times 10^{10}$; $f_{02} = 1{\cdot}080 \times 10^9$; $\lambda_{01} = 1060$ Å; $\lambda_{02} = 1750$ Å. Ramachandran has taken the value of $\chi_k = \chi_1 = \chi_2$ and has shown it to correspond with the lattice frequency 1332 cm^{-1}. Taking Krishnan's (1946 b) value of $\chi = 9{\cdot}6 \times 10^{-6}$ the calculated value of $\epsilon^{-1} d\epsilon/dT$ is found to be $+1{\cdot}25 \times 10^{-5}$ in very good agreement with the experimental value of $+1 \times 10^{-5}$.

ACKNOWLEDGMENT

The author's grateful thanks are due to Prof. R. S. Krishnan for his kind interest and valuable help in the progress of this work.

REFERENCES

BRETSCHER, E., 1934, *Trans. Faraday Soc.*, **30,** 684.
EUCKEN, A., and BÜCHNER, A., 1934, *Z. Phys. Chem.* B, **27,** 321.
FRÖHLICH, H., 1949, *Theory of Dielectrics* (Oxford : University Press), p. 109.
KRISHNAN, R. S., 1946 a, *Proc. Indian Acad. Sci.* A, **24,** 33; 1946 b, *Ibid.*, **24,** 45.
NARASIMHAN, P. T., 1953, *Proc. Indian Acad. Soc.* A, **37,** 551.
PETER, F., 1923, *Z. Phys.*, **15,** 358.
RAMACHANDRAN, G. N., 1947, *Proc. Indian Acad. Sci.* A, **25,** 266.
WATSON, H. E., RAO, G. G., and RAMASWAMY, K. L., 1934, *Proc. Roy. Soc.* A, **143, 558.**
WHITEHEAD, S., and HACKETT, W., 1939, *Proc. Phys. Soc.*, **51,** 173.

LETTERS TO THE EDITOR

Auger Effect in Semiconductors

It was suggested in 1950 by Fan and by Pincherle in unpublished reports that the Auger effect may be an important cause of recombination of holes and electrons in semiconductors, i.e. that the energy liberated when an electron and a hole recombine may be transferred to another free carrier. The importance of this effect, considered as the converse of impact ionization, is now recognized (see for example Burstein, Picus and Sclar 1955); its characteristic is that, when it operates alone, the lifetime of minority carriers is inversely proportional to the square of the density of the majority carriers. Such a dependence was demonstrated first by Moss (1953) in impure PbS and more recently by Haynes and Hornbeck (1955) in silicon.

The latter results will be compared with theory. As in this case the minority carriers recombine at deep traps, the theoretical treatment is simplified. Such traps may be thought of as foreign atoms, or ions, of radius R, only slightly perturbed by the surrounding periodic potential, and an atomic treatment may be used.

The probability per unit time that an Auger transition takes place is given by

$$W_{\mathrm{A}} = \frac{1}{\tau_{\mathrm{A}}} = \frac{2\pi}{\hbar}\,\rho(E)\left|\frac{e^2}{r_{12}}\right|^2$$

where τ_{A} is the lifetime of the minority carriers due to the Auger effect, E is the depth of the traps and $\rho(E)$ the density of states at energy E from the edge of the band occupied by the majority carriers. $\rho(E)$ will be taken as given by the standard expression

$$\rho(E) = \frac{4\pi V(2m)^{3/2}E^{1/2}}{h^3}, \quad \text{where } V = 4\pi R^3/3$$

$$\frac{1}{r_{12}} = \int \psi_{\mathrm{i}}^*(\mathbf{r}_1)\psi_{\mathrm{i}}^*(\mathbf{r}_2)\,\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\,\psi_{\mathrm{f}}(\mathbf{r}_1)\psi_{\mathrm{f}}(\mathbf{r}_2)\,d\mathbf{r}_1 d\mathbf{r}_2$$

(neglecting exchange). The suffixes i and f refer to the initial and final states.

If one assumes that for all particles $\mathbf{k}R \ll 1$ (where $\mathbf{k}$ is the crystal momentum) and takes spherically symmetrical wave functions, one finds

$$\tau_{\mathrm{A}} = \frac{C\hbar^4}{m^{3/2}e^4E^{1/2}R^7p^2},$$

p being the density of majority carriers. The numerical coefficient C depends on the particular form chosen for the wave functions, and with the above hypotheses is of order 10^{-2}. It is one or two orders of magnitude larger with p-type wave functions, depending on the k values.

For the case under consideration $E = 0{\cdot}8$ ev $= 1{\cdot}3 \times 10^{-12}$ erg; thus, inserting numerical values, $\tau_{\mathrm{A}}p^2 = 8 \times 10^{-25}\, CR^{-7}$ sec cm^{-6}. The biggest uncertainty is in the value of R, as one does not know the nature of the foreign atom. From the values of the most common ionic or covalent radii, one may expect R to be included between $0{\cdot}7$ and $1{\cdot}4 \times 10^{-8}$ cm. Thus considering also the range of variation of C, one finds the limits $10^{29} < \tau_{\mathrm{A}}p^2 < 10^{33}$ sec cm^{-6}. The experimental values are 10^{32} for n-type and 10^{34} for p-type silicon. The agreement may be considered satisfactory in view of the many uncertainties in the theory. The effective mass *m* of the ejected particle may also differ from the free electron mass and constitutes another unkown factor.

If recombination does not occur at a deep trap, but at a shallow one, or at a normal atom, the Auger lifetime is predicted to be smaller. This, according to Moss' results, would seem to occur in PbS.

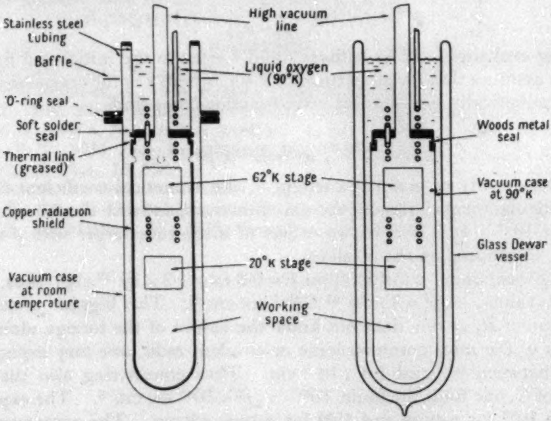
Thanks are extended to Dr. J. M. Radcliffe for profitable discussions. Acknowledgment is made to the Chief Scientist, Ministry of Supply and the Controller, H.M. Stationery Office, for permission to publish this letter.

Radar Research Establishment,
Malvern, Worcs.
L. Pincherle.
31st January 1955.

Burstein, E., Picus, G. S., and Sclar, N., 1955, *Proc. of the Atlantic City Conference on Photoconductivity*, in the press.
Haynes, J. R., and Hornbeck, J. A., 1955, *Phys. Rev.*, **97**, 311.
Moss, T. S., 1953, *Proc. Phys. Soc.* B, **66**, 993.

An Improved Miniature Liquefier-Cryostat

Miniature liquefier-cryostats in which the experimental working space is contained in the liquefier have a number of special advantages for work at liquid hydrogen and liquid helium temperatures. In the author's laboratory several such cryostats are in use of the type described by Chester and Jones. An improvement in design has recently been made in which the glass Dewar vessel containing the pre-coolant (liquid oxygen) surrounding the main vacuum case is replaced by a metal vessel of re-entrant type, which now serves also as the vacuum case. This is illustrated schematically in the figure for the case of a liquid hydrogen cryostat.


Stainless steel tubing
Baffle
'O'-ring seal
Soft solder seal
Thermal link (greased)
Copper radiation shield
Vacuum case at room temperature
High vacuum line
Liquid oxygen (90°K)
62°K stage
20°K stage
Working space
Woods metal seal
Vacuum case at 90°K
Glass Dewar vessel

New Type Old Type

By making the upper portion of this vessel of stainless steel tubing and providing a baffle to disperse the stream of evaporating oxygen, the temperature of the outer wall may be maintained only slightly below room temperature. The vacuum case can then be broken at an ‘ O ’-ring seal at a level *above* that of the top-plate of the cryostat, where a semi-permanent soft-soldered ring-joint is made as the last operation during construction. (As before, all other vacuum joints are hard-soldered.) The low-temperature part of the apparatus is protected by a copper radiation shield thermally linked to the top plate at 90°K by means of a greased screw-thread, and bearing a number of holes near this level to assist rapid evacuation. Under normal conditions we have noted a temperature difference between the bottom and top of the shield (of length about 30 cm) of about 30 degrees. This is no inconvenience, since the working space can in practice be protected by further shielding where appropriate to the experiment.

The new design leads to very greatly increased flexibility and accessibility. Limitations of space more or less disappear because the main vacuum case can now be made of any size or shape. It can be much more easily removed and also, being at room temperature, it can be fitted with additional openings or ‘ windows ’, or with appendages which may be separately detachable. For example, we have used a ‘ tail ’ of small diameter, attached via an ‘ O ’-ring seal. In this application the radiation shield also can be fitted with an easily detachable tail, thermally linked with the main shield via a greased screw thread A minor difference during operation follows from the fact that it is not possible to use ‘ exchange gas ’ to assist in cooling the contents to the temperature of liquid oxygen, but the only effect of this is to add a few minutes to the starting time. The over-all consumption of liquid oxygen is appreciably reduced by the new design.

I wish to acknowledge the help of Mr. W. A. G. Baldock and Mr. C. D. Murray in constructing and testing the new cryostat, and the support of the Department of Scientific and Industrial Research through a grant-in-aid.

Queen Mary College,
(University of London),
London, E.1.
30th March 1955.

G. O. Jones.

Chester, P. F., and Jones, G. O., 1953, *Proc. Phys. Soc.* B, **66**, 296.

REVIEWS OF BOOKS

Table of Sine and Cosine Integrals for Arguments from 10 *to* 100. U.S. Department of Commerce, N.B.S. Applied Mathematics Series 32 (Reissue of Mathematical Table 13). Pp. xv + 187. (Washington: U.S. Government Printing Office, 1954.) $ 2.25.

This is a re-issue with minor modifications and additions of a table issued in 1942 by the Mathematical Tables Project. The interval is 0·01 in the argument, and entries are given to 10 places of decimals. Second differences are printed in the table, and an ancillary table of coefficients is provided to facilitate the use of Everett's interpolation formula. H. H. HOPKINS.

A Concise History of Mathematics, by D. J. STRUIK. Pp. xix + 299. (London: Bell, 1954.) 14*s*.

Professor Struik has attempted to summarize the development of the main ideas in mathematics prior to 1900. The result is a very readable account of the subject for those who know rather more about mathematics than he presupposes. A particularly useful feature of the book is the inclusion of a good bibliography at the end of each chapter. H. H. HOPKINS.

Meteor Astronomy, by A. C. B. LOVELL. Pp. xiv + 463. (Oxford: Clarendon Press, 1954). 60*s*.

During the past quarter of a century a considerable effort has been devoted to the study of meteors. The development of radio techniques in recent years has enabled many of the outstanding problems to be solved so that the present time is appropriate for a comprehensive survey of the subject; and as the number of original papers that have been published is very large and as some of the more important of these are not readily accessible the need for such a survey has been much felt. This need is met by the admirable book which Professor A. C. B. Lovell has written.

In the early chapters the main observational methods (visual and photographic as well as radio) are lucidly described. The diurnal, seasonal, number and mass distribution of sporadic meteors are next discussed in detail, and a critical account is given of the measurements that have been made on their velocities, it being concluded that the percentage having hyperbolic velocities is very low and that the few cases of small hyperbolic velocities which appear to exist can probably be explained as arising from planetary perturbations. Most of the remainder of the volume concerns meteor showers. All the major showers are treated separately. The history of each is summarized and data given on the activity, radiant, velocities and orbit.

A vast amount of information is collected. It is presented clearly—mainly in the 175 tables and the 187 diagrams. The main text was completed in the early part of 1952 but an Appendix has been added in which results obtained up to the end of 1953 are included.

Professor Lovell is to be warmly congratulated on *Meteor Astronomy* which is certain to become a standard source of reference.

The physics of meteors (their evaporation while passing through the Earth's atmosphere, the luminosity and ionization they produce, etc.) is reserved for a later book. **D. R. BATES.**

BOOK NOTICES

Bibilography of Books and Published Reports on Gas Turbines, Jet Propulsion and Rocket Power Plants, January through December, 1953. Pp. iv + 110. (Washington : United States Department of Commerce, National Bureau of Standards, Supplement to NBS Circular 509, 1954.)

Statistical Theory of Extreme Values and some Practical Applications (A Series of Lectures), by E. J. GUMBEL. Pp. viii + 51. (Washington : United States Department of Commerce, National Bureau of Standards, Applied Mathematics Series 33, 1954.)

Reviews of Research on Problems of Utilization of Saline Water. Pp. 96. (Paris : Unesco, Arid Zone Programme–IV, 1954.)

Chromatography (British Medical Bulletin, Vol. 10, No. 3, 1954). Pp. xi + 252. (London : The Medical Department, The British Council.) 15*s*.

CONTENTS OF SECTION A

PAGE

Mr. J. M. Dickson, Mr. B. Rose and Mr. D. C. Salter. Elastic Scattering of Polarized Protons from Complex Nuclei 361

Dr. P. J. Grant, Dr. J. G. Rutherglen, Dr. F. C. Flack and Dr. G. W. Hutchinson. Angular Distributions and Angular Correlations in the Reaction ^{23}Na(p, γ)^{24}Mg 369

Dr. U. Öpik. The Layzer Approximation for the Treatment of Collisions of Electrons with Atoms 377

Dr. L. L. Green, Mr. J. P. Scanlon and Dr. J. C. Willmott. Angular Distributions from (d, n) Reactions at 0·86 mev Deuteron Energy 386

Mr. M. A. Ihsan. An Investigation of ^{7}Li(d, n)^{8}Be and ^{11}B(d, n)^{12}C Nuclear Reactions 393

Dr. J. H. Fremlin and Miss V. M. Spiers. The Break-up of Negative Ions in the Cyclotron. 398

Dr. D. T. Stewart. Electron Excitation Functions of Infra-Red Nitrogen Spectra 404

Dr. B. G. Owen and Prof. J. G. Wilson. Measurements of the Momentum Spectrum of μ-Mesons at Sea Level: I—The Momentum Range 5×10^{8} ev/c–2×10^{10} ev/c 409

Dr. L. Cohen. A Variation-Perturbation Method: I 419

Dr. L. Cohen. A Variation-Perturbation Method: II 425

Dr. W. R. Hindmarsh and Dr. H. Kuhn. Isotope Shifts and Hyperfine Structure in the Atomic Spectrum of Tin 433

Research Notes:

Mr. Amalkumar Raychaudhuri. Note on the Network Approximation in Metals 439

Mr. J. Hubbard. On the Interaction of Electrons in Metals. 441

Dr. M. A. Jaswon and Mr. H. Aghajanian. The Interaction of Two Excited Hydrogen Atoms 444

Dr. D. S. Schonland. The Use of Overlapping Intervals in Multiple Scattering Measurements 447

Letters to the Editor:

Mr. G. J. Ritter and Dr. G. W. Series. Nuclear Electric Quadrupole Moment of Potassium 39 450

Mr. R. Batchelor. The Yield of ^{7}Li(p, n)^{7}Be* Neutrons for Proton Energies between 2·5 and 2·9 mev. 452

Reviews of Books 454

Contents of Section B 456

The Scattering of Light in Sodium Chloride Monocrystals

By S. P. F. HUMPHREYS-OWEN
Birkbeck College, University of London

MS. received 15*th November* 1954 *and in amended form* 24*th January* 1955

Abstract. The total intensity of light scattered at 90° by specimens of natural and synthetic NaCl single crystals has been investigated as a function of wavelength from 6600 Å to 2500 Å. The scattered intensity is found to be of an order 40 times greater than that expected from thermal theory, and the wavelength function is not inverse fourth power throughout the spectral region studied.

Therefore only a small part of the scattering can be attributed to thermal fluctuations and the bulk is attributed to a submicroscopic block structure of the crystals. From the wavelength function an estimate has been made of the linear dimension of the blocks, and this is found to differ little from specimen to specimen and from place to place in one specimen; it is of the order 1500 Å. Thus the above-mentioned structure seems to be an intrinsic feature of the real structure of single crystals.

§ 1. Introduction

The order and the closeness of packing of the atomic scattering centres of a perfect crystal at 0°K would result in zero scattering of light of optical wavelengths because of destructive interference. At higher temperatures, however, sound waves of thermal origin traverse the crystal in all directions, causing stratifications of density which give rise to a type of scattering called thermal. The intensity of this scattered light is inversely proportional to the fourth power of the wavelength of the irradiating light (hereafter referred to as the λ^{-4} law), and directly to the absolute temperature.

A real crystal would be expected to yield an additional scattered intensity, superposed on the thermal scattering, by reason of imperfections which act as locations of optical discontinuity and modify the total optical interference. Experimenters in the past have sought to isolate the thermal fraction of the scattered light on the assumption that imperfections are accidental and not intrinsic to the crystal, and not therefore a significant source of scattering. For example Krishnan and Chandrasekharan (1950) concentrate on the wavelength-shifted Brillouin components peculiar to thermal scattering, and Landsberg and Mandelstam (1932) confine attention to the temperature-dependent fraction on the assumption that only thermal scattering is a function of temperature. But this expedient will fail if there is another source of scattering, itself temperature-dependent.

The present study was suggested to the author by Dr. R. Fürth who (1949) put forward reasons for supposing that the real structure of a crystal, apart from accidental dislocation, consists of a mosaic of perfect blocks located in a three-dimensional net of disordered material. The existence of this 'intrinsic block structure' had first been proposed, on theoretical grounds, by Born (1947), who in a discussion of the consequences of the anharmonic character of the restoring forces in a perfect lattice (also responsible for the thermal expansion) was led to the consideration of a certain length l which he expressed by

$$l=\frac{2a}{\beta\theta}\tanh\frac{\theta}{2T}, \qquad \ldots\ldots(1)$$

where a is the lattice constant, β the coefficient of linear expansion, θ the Debye temperature of the substance, and T the Kelvin temperature. This expression yields values for l of the order 1000 Å at room temperature for most materials.

This length roughly represented a limit to the validity of the method of approximation employed, but Born, at the time, thought it probable that in addition it might represent a real limit to the extent of the perfect lattice. In later publications (Born 1951) he expressed doubts as to this view. I understand, however, from Dr. Fürth that the latter has in the meantime been able to give a more convincing proof of the existence of a real critical length on the basis of rigorous lattice theory, which he hopes to publish in the near furture.

A system of this kind will have a finite scattering power because the net of disorder around the perfect blocks provides the necessary optical discontinuity, just as transparent particles in a transparent medium will scatter light if there is a relative refractive index different from unity. Moreover, the scattered intensity will have characteristics provided by the shape and size of the blocks, and in particular will not obey the λ^{-4} law if irradiated by light of wavelength comparable with 1000 Å; for the λ^{-4} law ceases to hold when the linear dimension of the scattering particle becomes greater than about one tenth of the wavelength.

Thus an investigation of the scattered intensity as a function of wavelength as far as possible into the ultra-violet should yield information by virtue of departure from the λ^{-4} law; and furthermore, if the experimental wavelength function contains features common to many crystal specimens, this will be evidence for something intrinsic and not accidental. The method by which an estimate of the size of the crystal scatterers has been obtained from the experimental wavelength functions is outlined in the next section.

§ 2. A MODEL FOR THE INTERPRETATION OF THE WAVELENGTH FUNCTIONS

Let it be assumed that the blocks are of irregular compact shape of linear dimension r, whose lattice in their interior is perfect, but whose boundaries are defined by a thin (of atomic dimension) net of disordered atoms. The scattered light is then entirely due to the superposition of the wavelets issuing from the disordered atoms, and hence, to a first approximation, should have the same angular and spectral distribution of intensity as that due to an assembly of spherical particles of radius r and refractive index equal to that of the crystal. The scattering behaviour of isolated spherical particles has been fully computed from the well-known theory of Mie (1908). One cannot, however, apply this theory to the present model of close-packed blocks without further justification. If the close-packed blocks were all exactly alike and of regular geometrical form the light scattered by many block boundaries would possess definite phase relationships and would sum to a total effect which would be different from that obtained from randomly distributed blocks. But our assumption is that the blocks are irregular in shape and not all exactly alike. In this case it is permissible to argue that the regularities of phase of light scattered from many boundaries will be largely destroyed, and that the total effect will approximate to that of blocks of similar, regular, shape, but randomly distributed, thus enabling the Mie theory to be employed.

Let the incident light be of wavelength λ and be unpolarized, and let i be the total (sum of polarized components) flux scattered per unit solid angle at 90° to the incident beam, per unit incident intensity per unit volume of scattering material. Let the scattering system be a random assembly *in vacuo* of similar dielectric spheres of radius r. For them, the function i will be made up of two factors, μ and ν say: μ almost independent of wavelength and a function of the number of spheres per unit volume and of their refractive index, and ν which is a function of the parameter $2\pi r/\lambda$. There is no reason to expect a direct relationship between the function μ for the assembly of blocks and the corresponding function for the assembly of spheres as long as the number of dislocated atoms on the block boundaries is not specified. But, on the other hand, it is to be expected that the function ν should be approximately the same in both cases. Thus, although no information about the block structure can be obtained by measuring the absolute magnitude of i at constant λ, an investigation of i as a function of λ will give information on r from the behaviour of the function $\nu(r, \lambda)$, using the sphere model.

When, from tables of the Mie theory (see references), for spheres of given r, the function i is plotted against λ (the vacuum wavelength in this model), and other variables are held constant, there are two salient features in the curve:

(i) there is a wavelength λ_1 say, below which the power n in $i \propto \lambda^{-n}$ begins to fall below 4 (with accuracy to 1%), and

$$2r = 0{\cdot}8\lambda_1/\pi, \qquad \ldots\ldots(2)$$

(ii) there is a shorter wavelength λ_2 say, at which the curve passes through a maximum, and

$$2r = 1{\cdot}8\lambda_2/\pi. \qquad \ldots\ldots(3)$$

In order to evaluate the equivalent diameter $2r$ of the crystal blocks either or both of these features may be employed, but if both can be observed experimentally, the confirmation of the relation

$$\lambda_1/\lambda_2 = 1{\cdot}8/0{\cdot}8 = 2{\cdot}25 \qquad \ldots\ldots(4)$$

will lend confidence to the underlying assumption that the blocks approximate in their behaviour to spheres.

A third test may be tentatively applied: if $2\pi r/\lambda$ is small, the scattering of a sphere at constant λ is proportional to r^3, other variables being held constant. If therefore i, at a fixed long wavelength, is compared for different crystals or different parts of one crystal, we might ascribe differences in i to differing equivalent radius if $i \propto r^3$ is found to hold (r having previously been estimated). But in view of the arbitrariness in the number of dislocated atoms emphasized above, it will be no loss to the model if this is found not to be the case, for the individual history of each crystal would be expected to have an unpredictable effect on the number of dislocated atoms.

§ 3. EXPERIMENTAL

The study of the wavelength variation of i consists of measuring a ratio of scattered flux to incident intensity under conditions in which λ is the only variable. The experimental arrangement is shown schematically in figure 1. The source of light was a Siemens xenon gas arc type E.2, which provides a continuum whose energy output is level from 6800 Å to 3800 Å, and then falls to a fifth at 2500 Å. This source was found preferable to a line spectrum. After monochromation (by a Hilger Barfit wavelength spectrometer with adaptor D.222) the light was modulated to provide a frequency of 22 c/s; a 44 c/s oscillator drove a Ferranti synchronous motor which (designed for 375 rev min^{-1} at 50 c/s) rotated at 330 rev min^{-1}. A four-sector chopper thus yielded a square wave of 22 c/s. This frequency avoids beating between a Fourier component of the square wave and the 100 c/s modulation of a light beam whose source is run from 50 c/s mains.

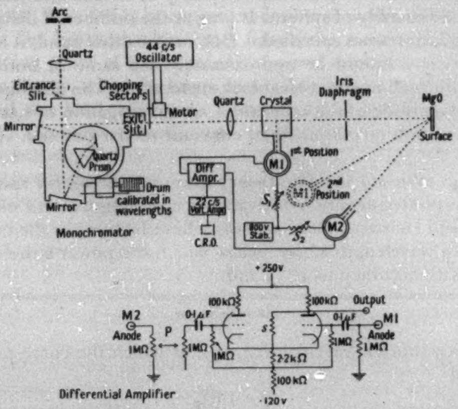

Arc
Quartz
44 c/s Oscillator
Entrance Slit
Chopping Sectors
Motor
Quartz
Crystal
Iris Diaphragm
MgO
Surface
Mirror
Exit Slit
Quartz Prism
1st Position
M1
Diff Ampr.
2nd M1 Position
S_1
Drum calibrated in wavelengths
22 c/s Volt Ampr
Mirror
800 V Stab.
S_2
M2
Monochromator
C.R.O.
+250 V
100 kΩ
100 kΩ
Output
M2
0·1 μF
0·1 μF
M1
Anode
P
s
Anode
1 MΩ
1 MΩ
1 MΩ
1 MΩ
1 MΩ
1 MΩ
2·2 kΩ
100 kΩ
Differential Amplifier
-120 V

Figure 1. Schematic experimental arrangement, and circuit of differential amplifier for comparing outputs from two photomultipliers.

The beam was focused on to the centre of a crystal specimen in the form of a cube block of about 2 cm side. All faces were highly polished, and then the three not wanted for the transmission of light were painted black. Scattered light was detected by a photomultiplier (M1 in the figure) of RCA type 1P.28 which has a special glass envelope transmitting to about 2400 Å. In principle, the ratio is required, wavelength by wavelength, of the resulting photocurrent to that produced when M1 views the incident beam directly. Photomultiplier outputs are, however, linear with luminous flux only for low flux levels, and therefore, in a second experiment, M1 ('second position' in the figure) received a small fraction of the flux diffusely reflected from a magnesium oxide surface. After a small correction for the wavelength variation of the reflectance of MgO (Menzies and Skinner 1949) the response of M1 in the second position is proportional to the intensity incident in the crystal. Experiment showed that the difference in air path to M1 between the two positions introduced no detectable new wavelength variable, at any rate as far as 2500 Å. As regards the path in the crystal, it will be seen that, in each case, the same path length and transmission through two crystal surfaces is involved.

In order, however, to eliminate errors due to changes in light output and electronic characteristics occurring in the course of the two experiments, a null method was devised in which, in each experiment, a fraction of the output of a second mutiplier, M2, viewing the MgO surface, was made equal to that of M1. The position of M2 was never changed. At a given wavelength, let this fraction be p_1 in the first (scattering) position of M1, then the outputs are in the ratio

$$\mathrm{M1\,(scattering)/M2} = p_1.$$

Similarly, when for M1 in the second (incident) position, equality is obtained at a fraction p_2 of M2's output, we have

$$\mathrm{M1\,(incident)/M2} = p_2.$$

Thus the ratio p_1/p_2 eliminates the characteristics of M2, which acts merely as a transfer instrument, and this ratio, which is proportional to the desired function i, is studied as a function of λ. Of course it is not necessary to use a photomultiplier for M2; any ultra-violet sensitive cell would serve. But another 1P.28 happened to be available, and in any case for the null technique described next it was desirable to have two cells of similar electrical characteristics.

In the differential amplifier of figure 1 equal signals at the grids produce zero output apart from a small residue proportional to the sum of the two signals. This residue can be reduced to a negligible value by choice of the resistance s (McFee 1950) which is fixed empirically and is of the order 150 kΩ in the circuit shown. Since an a.c. null is required phasing must be correct and in practice it was found sufficient to make components on each side of the amplifier equal, to 1% tolerance. The output of M2 was controlled by the constant impedance potentiometer P (two 1 megohm carbon potentiometers ganged), and the fraction p was calibrated directly on a dial. The double triode was type 6SL7.

The off-null output from the differential amplifier was further amplified in a narrow bandwidth voltage amplifier and finally displayed as a sinusoidal 22 c/s signal on an oscilloscope. The scattered flux from good crystal specimens is very small indeed, especially after monochromation and in the ultra-violet where the output from the arc falls off. Even with an amplifier bandwidth as low as a fraction of a cycle some residual noise remains, and it was found that the cathode-ray oscilloscope display was the best way of seeing the signal through the noise.

The resistances S_1 and S_2 in the respective lines to the potential dividers supplying the multiplier dynodes were variable and served the purpose of producing a null at a convenient initial setting of the potentiometer P. In order, however, to preserve the purely transfer function of M2, the setting of S_2 was not altered once it had been set in the first (position 1) part of an experiment.

§ 4. RESULTS AND DISCUSSION

The onus of accuracy falls on the linearity of the multiplier M1, which is required in this work to detect over the full range of its spectral sensitivity. In a preliminary note (Fürth and Humphreys-Owen 1951) certain kinks in the curve of i plotted against λ were recorded which have since been found to be spurious and due to non-linearity in a multiplier. The present communication should therefore be regarded as superseding the above report.

Scattered light entered the multiplier from a region in the crystal about 5 mm long and 1 mm^2 across; this is small enough for different parts of one crystal to be studied. Results for five crystals are presented, three natural and two synthetic, of which one was grown from melt (method of Kyropoulos by Messrs. Hilger and Watts) and one from solution (by Prof. Chapelle of the University of Nancy). Except for the latter, which had a higher scattering power than the others, the monochromatic beam in the crystal was almost invisible to the naked eye when viewed at 90°, but could be seen when viewed from the forward direction. This is a first indication that there are scatterers in the crystals which are not small compared with the wavelength, for the scattering from small particles, small enough to yield the λ^{-4} law, is symmetrical about 90° to the incident beam. Before any experiment, care was taken to view the beam and ensure that no bright spots indicating gross imperfections were in it. The general appearance of the beam to the eye was recorded in each case, viewed from the forward direction. The three natural crystals and the synthetic crystal from the melt presented a similar beam, not uniform in brightness, with the 'milky' smoothness of a beam in dust-free liquid, but slightly 'spotty'. In one natural crystal it was possible to select two regions for study, one with beam apparently almost free of these discrete spots and the other with considerable concentrations of them; it should, however, be noted that the spots were only slightly brighter than their background. The synthetic crystal grown from solution, on the other hand, had a much brighter beam than the others, bright enough to be seen at 90°, and it was apparently uniform in brightness. This crystal was peculiar in its history: it had been ground in the form of a cylinder, for the purpose of angle measurements in France, and then annealed and polished. None of the other crystals was annealed.

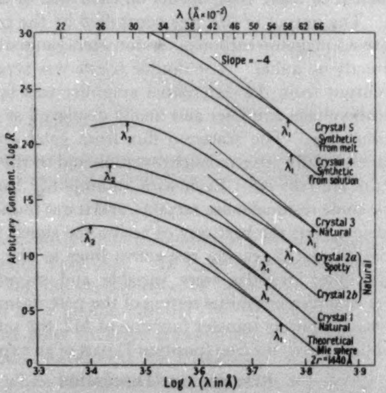


Figure 2. Logarithmic curves from which scatterer size is estimated from wavelength of departure λ_1 from λ^{-4} law, and from turning points (λ_2).

The presentation is in the form of the ratio $R = p_1/p_2$ as a function of the vacuum wavelength λ. In order to show the behaviour of the power n in $R \propto \lambda^{-n}$, log R has been plotted against log λ (figure 2).

The absolute magnitude of R did not differ sufficiently from crystal to crystal for the curves to be conveniently drawn in correct relative location on the axis of $\log R$, therefore comparison between crystals should be made of their slopes only. The comparison of magnitudes at fixed wavelength is presented in the table (see below).

It is seen that in no case is the λ^{-4} law obeyed through the spectral region studied, but there is initial obedience at long wavelengths, and to show this, the straight line of slope -4 has been drawn in each case. The wavelength λ_1 of departure from the law is indicated, and also the wavelength λ_2 at which the curve turns over, in the two instances where this was observed. Neither of the synthetic crystals could be studied further into the ultra-violet than shown, because of fluorescence.

Two curves for one crystal (crystal 2) are shown, one for a region in which the beam was clear, and the other for a region in which it was 'spotty'. At the bottom of the figure is a theoretical curve for a system of spheres of equal radius.

Certain control experiments were carried out, and are illustrated by the curves of figure 3. These are of R plotted against λ^{-4}, and show actual experimental points in order to indicate the degree of random error in the technique. The curves are: *a* a typical curve for crystal 3, *b* and *c* respectively curves for large imperfections in a block of fused silica and for a liquid (benzene). It is important to show that there is no appreciable spurious curvature introduced into scattering curves for any reason. Experience showed that a serious cause of such spurious effects was when the multiplier M1 was not operating on the same linear response curve throughout all experiments. It was therefore arranged that in all experiments M1 gave the same response at the wavelength of its peak response. From the fact that the correct curves, namely $i \neq f(\lambda)$ for the large imperfections, and $i \propto \lambda^{-4}$ for the liquid, have been obtained in these control experiments, we may have confidence that, in the other experiments also, spurious curvature is absent.

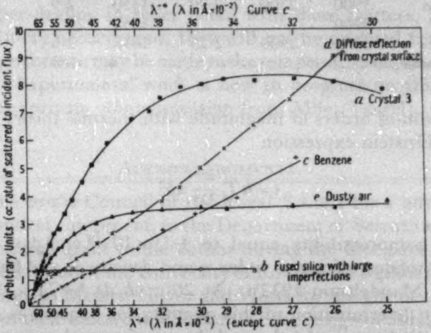


Figure 3. Scattering ratio plotted against λ^{-4}, for a crystal and for various other scatterers, in order to show absence of consistent error in the experimental technique.

Curve *d* is a curve for diffuse reflection from a crystal surface. Because of imperfect polish and slight moisture haze there is some diffuse reflection from the entrance and exit faces of the crystal, and this light is in its turn diffusely reflected at the face through which scattered light enters the multiplier. The expedient of immersing the crystal in a liquid of similar refractive index so as to reduce such reflections was not followed in this work, because of absorption in the ultra-violet. Hence some criterion was necessary that diffuse reflections were negligible. This was provided by the fact that the curve for diffuse reflection does not pass through the origin, whereas the curve for true crystal scattering does, and by the fact that the curve for diffuse reflection continues to rise in the ultra-violet, whereas the curve for crystal scattering decreases in slope.

Curve *e* is for dusty air in a closed cell after the heavier particles had been allowed to fall. This is to emphasize the different form of curves, on the one hand for a *distribution* of particle sizes (dust), and on the other hand for a system (the crystal) whose scattering elements appear to differ little in size.

The table gives the values of λ_1, λ_2, λ_1/λ_2, estimated $2r$, comparative scattered intensity at 6400 Å, comparative r^3 relative to the smallest block size, and comparative intensity for benzene.

Crystal	λ_1 (Å)	λ_2 (Å)	λ_1/λ_2	$2r$ (Å)	Int.	r^3
1	5300			1360	1·00	1·00
2*b*	5300			1360	2·6	1·0
5	5800			1490	2·0	1·3
2*a*	5900	2700	2·20	1550	2·9	1·5
4	5900			1550	20	1·5
3	6500	2900	2·25	1670	2·6	1·85
Benzene					5·0	

For comparison of orders of magnitude with thermal theory we may use the Smoluchowski–Einstein expression

$$i = \boldsymbol{k}T \frac{\pi^2}{2\lambda^4} x\delta^2 \qquad \ldots\ldots(5)$$

where x is the compressibility equal to $4{\cdot}23 \times 10^{-12}$ cm² dyn⁻¹, for rock-salt, and $\delta = 2n\rho\partial n/\partial\rho$ relates refractive index n and density ρ, and has the value 0·96 (Landsberg and Mandelstam 1932). At 20°c, $di/d\lambda^{-4} = 7{\cdot}8 \times 10^{-25}$ cm³.

Now a recent determination of this quantity for benzene is (Kushner 1954) $1{\cdot}66 \times 10^{-22}$ cm³, and so, on the thermal theory of crystal scattering, the ratio of intensities NaCl to benzene should be 0·0047. Actually it is 0·2 in the case of crystal 1, the weakest scatterer of this research, and thus the scattering is at least 40 times more intense than predicted on thermal theory.

We shall therefore ignore the thermal contribution to the scattering in discussing the results.

Referring to the curves of figure 2, we notice first the general similarity of their form to that for a theoretical distribution of spheres, and secondly the good correlation between λ_1 and λ_2 in the two relevant cases. These observations lend support to the assumption that the scattering system approximates to that of isolated spheres, and that, at least within the volume viewed in one experiment, the elements of the system are of the same size. The purpose of showing the scattering curve for airborne dust in figure 3 is to emphasize that the technique is sensitive to the distinction between a distribution of sizes and an assembly of elements of nearly equal size.

We may note in passing that another assumption, namely that the elements are compact in shape, is supported by the observations, and that therefore the suggestion (Taurel 1952) that non-thermal scattering is due to one-dimensional cracks seems difficult to accept.

Next, it can be noted that when different crystals, and different regions of one crystal, are studied, there is a remarkably small range of difference amongst the sizes inferred for the scattering elements. The correlation between size and absolute intensity of scattering is, as seen from the table, not good; but this correlation, as mentioned earlier, is not necessarily to be expected.

Now one would expect that accidental imperfections caused by non-ideal growing conditions or by later disturbances would have a wide distribution of size, and the contrast between this expectation and the uniformity of size actually observed gives one the right to suggest that something truly intrinsic is being observed.

It is perhaps too much to say that these results form a close verification of the views of Born and Fürth; their conclusions as so far published are still only as to order of magnitude. But nevertheless these views do predict a structure with properties which are in accordance, not only with the present experiments on light scattering, but also with other aspects of physical behaviour such as melting and tensile strength. Since I understand that Dr. Fürth has under preparation a paper dealing in fuller detail with these matters, as well as with comparison with x-ray observations, they will not be pursued further here.

In conclusion, reference may be made to the temperature dependence predicted by equation (1). Experimental work is now in progress on this aspect of the problem elsewhere (private communication from Mlle. Taurel).

ACKNOWLEDGMENTS

Thanks are due to the Council of the Royal Society for a government grant, to Dr. Fürth for optical equipment, to the Department of Scientific and Industrial Research for a research grant to the author during the first part of the work, to the Central Research Fund of the University of London for a grant for the purchase of the Siemens lamp, to Messrs. Hilger and Watts for the gift of crystal material, to Mr. P. Rippon of their Electronics Section for the design of the light modulation technique, and to Professor J. P. Chapelle and Mlle. L. Taurel of the University of Nancy for the loan of a crystal and for helpful discussion.

REFERENCES

BORN, M., 1947, *Proc. Math. Phys. Soc. Egypt*, **3**, 35 ; 1951, *Festschr. Akad. Wiss. Gött., Math.-phys.* Kl. I, No. 1, and *Nachr. Akad. Wiss. Göttingen*, Kl. IIa, No. 6.
FÜRTH, R., 1949, *Phil. Mag.*, **40**, 1227.
FÜRTH, R., and HUMPHREYS-OWEN, S. P. F., 1951, *Nature, Lond.*, **167**, 715.
KRISHNAN, R. S., and CHANDRASEKHARAN, V., 1950, *Proc. Indian Acad. Sci.*, **31**, 427.
KUSHNER, L. M., 1954, *J. Opt. Soc. Amer.*, **44**, 155.
LANDSBERG, G., and MANDELSTAM, L., 1932, *Z. Phys.*, **73**, 502.
McFEE, R., 1950, *Rev. Sci. Instrum.*, **21**, 770.
MENZIES, A. C., and SKINNER, J., 1949, *J. Sci. Instrum.*, **26**, 299.
MIE, G., 1908, *Ann. Phys.*, **25**, 377.
TAUREL, L., 1952, *C.R. Acad. Sci., Paris*, **234**, 2443.
U.S. NAT. BUR. STAND., 1949, *Appl. Maths Ser.* No. 4. (Tables of scattering functions.)

Experimental Investigation of the Motions of Electrons in a Gas in the Presence of a Magnetic Field

By B. I. H. HALL
Department of Physics, University of Adelaide, South Australia

Communicated by L. G. H. Huxley ;
MS. received 14*th July* 1954 *and in revised form* 23*rd December* 1954

Abstract. Results are given of experiments to measure the drift velocities of electrons diffusing in hydrogen under the influence of both electric and magnetic fields. Two cases are considered: (*a*) the magnetic field B in the same direction as the electric field Z, (*b*) B perpendicular to Z. The experimental results are interpreted in terms of a convenient microscopic model.

---

§ 1. Introduction

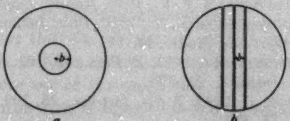
The theory of the spread of a stream of electrons from a small source in a gas in the presence of a magnetic field has been given previously (Huxley 1940, Huxley and Zaazou 1949).

In the present experiments the electrons enter the diffusion chamber through a small central hole in the cathode and move through the gas under the action of a constant and uniform electric field. The anode is a plane electrode partitioned into sections appropriate to the particular measurements. The spread of the stream and the characteristics of the motion are found by measuring the ratios of the individual currents to the constituent sections of the anode. Hydrogen was the gas chosen because of its general suitability.

The aim of the present investigation was to show that the theory of this method as formulated by Huxley (1940, 1949) describes correctly the spread of the electron stream. It is also shown that by adopting a well-known kinematic model (Appendix I) describing the collisions between the electrons and the gas molecules consistent values for the various microscopic parameters of the electron motion can be obtained.

§ 2. Diffusion without a Magnetic Field

Two forms of division of the anode have been employed: (*a*) into annuli, (*b*) into strips (figure 1).



a Annular Electrode

b Strip Electrode

Figure 1. Annular and strip electrodes.

(*a*) *Division into annuli.*

It has been shown (Huxley 1940) that

$$R_b = 1 - \frac{\exp[-\lambda h\{(1+b^2/h^2)^{1/2}-1\}]}{(1+b^2/h^2)^{1/2}} \qquad \ldots\ldots(1)$$

where R_b is the ratio of the current falling on an annulus of radius b to the total current entering the chamber, and $\lambda h = 20{\cdot}15Zh/k_1$ where h is the length of the diffusion chamber ($=2$ cm), and Z is the electric field strength in v cm^{-1}.

This expression has been thoroughly tested (Crompton and Sutton 1952, Crompton, Huxley and Sutton 1953) and as shown in a recent note by Huxley and Crompton (1955) describes accurately and uniquely the distribution of electrons in the stream.

(*b*) *Division into strips.*

The ratio R_b of the current falling on the central strip of width $2b$ to the total current is (Huxley 1940)

$$R_b = \int_0^b \frac{K_1[\lambda(x^2+h^2)^{1/2}]dx}{(x^2+h^2)^{1/2}} \Big/ \int_0^\infty \frac{K_1[\lambda(x^2+h^2)^{1/2}]dx}{(x^2+h^2)^{1/2}}$$

$$= \left(\frac{2}{\pi}\right)(\lambda h e^{\lambda h}) \int_0^{b/h} \frac{K_1[\lambda h(1+s^2)^{1/2}]ds}{(1+s^2)^{1/2}}. \qquad \ldots\ldots(2)\dagger$$

In the apparatus used $b/h = 0{\cdot}25$ and the value of the integral was computed (using Simpson's rule) for varying values of λ. R_b is shown as a function of $Zh/k_1(=hW/40{\cdot}3K)$ in figure 2; k_1 is proportional to Townsend's energy factor, W is the electronic drift velocity and K the coefficient of diffusion.

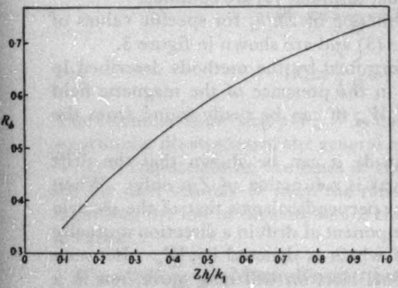


Figure 2. Ratio curve for strip electrode with $b/h = 0{\cdot}25$.

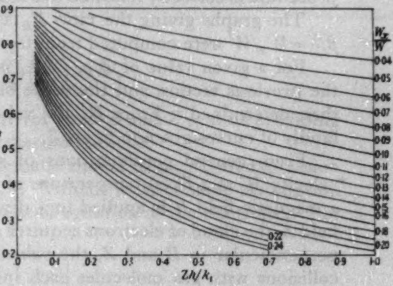


Figure 3. Ratio curves for W_x/W = constant.

The ratio R_b is measured for a chosen value of the electric field Z and the pressure p, and Zh/k_1, and hence k_1, found from the curve. As for the case when the anode is divided into annuli the agreement between the values of k_1 measured at different pressures for the same value of the ratio Z/p is excellent, and although the geometry differs markedly in the two cases, the values of k_1 are practically identical as can be seen from the examples given in table 1.

---

† K_1 is a modified Bessel function.

Table 1

Z/p (v cm^{-1} mm^{-1})	0·2	0·5	1·0
k_1 (strip electrode)	2·59	5·17	9·10
k_1 (annular electrode)	2·55	5·12	9·04

The vacuum system, electrometers and current measuring devices used were the same as those employed by Crompton and Sutton (1952), and as in their experiments the source of electrons was an uncoated platinum filament.

§ 3. Diffusion and Drift in a Transverse Magnetic Field

In this case the magnetic field B is perpendicular to the electric field Z.

(*a*) Huxley and Zaazou (1949) obtained the following expressions for the ratio R_B of the current falling on the halves of the anode when the field B is parallel to the central slit (figure 1 (*b*)).

$$\left.\begin{aligned} R_B &= \frac{I_1 - I_2}{I_1 + I_2} \\ \text{where} \qquad I_1 &= \frac{\pi}{2\lambda h}\exp\{-\lambda h(1-\beta^2/\lambda^2)^{1/2}\} \\ \text{and} \qquad I_2 &= \sum_{n=0}^{\infty}\frac{(\beta/\lambda)^{2n+1}(\lambda h)^n}{1\,.\,3\,.\,5\ldots(2n+1)}K_n(\lambda h) \\ 2\beta &= W_x/K_B. \end{aligned}\right\} \qquad \ldots\ldots(3)$$

These formulae are obtained from the differential equation for the electron concentration given in Appendix II. Although they are correct they were originally deduced on the hypothesis that the point source behaved as a dipole source. However experiment shows, as described in the note by Huxley and Crompton, that the point source acts as a pole source and when the mathematical procedure is correctly followed it is found that formulae (3) are obtained.

The graphs giving the ratio R_B as a function of Zh/k_1 for specific values of $\beta/\lambda = W_x/W$ were computed from formulae (3) and are shown in figure 3.

For a given value of Z/p, Zh/k_1 is determined by the methods described in the previous section and the value of R_B in the presence of the magnetic field then determined. From R_B the value of W_x/W can be easily found from the family of curves given in figure 3.

From general considerations of similitude it can be shown that the drift velocity W at a fixed temperature of the gas is a function of Z/p only. When a magnetic field B is applied in a direction perpendicular to that of the electric field Z, the cloud of electrons acquires a component of drift in a direction mutually perpendicular to B and Z, the velocity of which is denoted by W_x. Between collisions with gas molecules each individual electron will now move, not in a straight line, but along a curved path which is an element of a trochoidal helix. However, if the field B is small and the pressure p at least of the order of two millimetres of mercury, these elementary paths may be considered as approximately straight. In that event, for constant B and Z/p, it is easily seen that W_x will be inversely proportional to the pressure p of the gas, that is to say the product pW_x/W should be found to be constant.

The results in table 2 support this expectation and indicate that formulae (3) together with the macroscopic assumptions upon which they are based suffice to describe correctly the distribution of electrons in the stream under the conditions which apply in the present experiments.

Table 2. Table of Values of pW_x/W when $B=21{\cdot}5$ gauss

Z/p	$p=2$ mm	3 mm	4 mm	5 mm	6 mm
0·1				0·660	0·654
0·2			0·520	0·515	0·510
0·3	0·440	0·432	0·432	0·445	
0·4	0·382	0·390	0·382	0·385	
0·5	0·336	0·334	0·336	0·335	
0·6	0·302	0·300	0·300		
0·7	0·278	0·276	0·276		
0·8	0·254	0·252	0·256		
0·9	0·234	0·234	0·232		
1·0	0·224	0·226	0·224		
1·5	0·184	0·183	0·184		
2·0	0·162	0·159			
3·0	0·140	0·142			

(*b*) As yet no assumptions have been made concerning the nature of the collisions between the electrons and the gas molecules, but on the basis of the model as described in Appendix I, it has been shown by Huxley (1937, 1951) that

$$\left.\begin{aligned} W_x &= \frac{2}{3}\frac{Z}{B}\overline{\left\{\frac{\omega^2T^2}{1+\omega^2T^2}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right]\right\}} \\ \text{and}\qquad W_z &= \frac{2}{3}\frac{Z}{B}\overline{\left\{\frac{\omega T}{1+\omega^2T^2}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right]\right\}} \end{aligned}\right\} \qquad \ldots\ldots(4)$$

where $\omega=Be/m$ and $T=l/U=$ the time between collisions for an electron moving with speed U along a free path of length l.

In the present experiments $\omega^2T^2 \ll 1$, whence

$$W_x \simeq \frac{2}{3}\frac{Z}{B}\omega^2\overline{T^2}$$

and

$$W_z \simeq \frac{2}{3}\frac{Z}{B}\omega\overline{T} = \frac{2}{3}\frac{Ze}{m}\overline{\left(\frac{l}{U}\right)} = W.$$

The formulae show that pW_x/W is constant for fixed B and Z/p, thus providing a particular illustration of the general result discussed above.

It is impossible on the basis of the kinematic model to derive an expression for the drift velocity W from the ratio W_x/W without making some assumption regarding the law of distribution of the speeds of agitation of the electrons. For instance

$$\left.\begin{aligned} W_M &= 0{\cdot}85\,\frac{Z}{B}\frac{W_x}{W}\ \text{(Maxwell's distribution)} \\ W_D &= 0{\cdot}943\,\frac{Z}{B}\frac{W_x}{W}\ \text{(Druyvesteyn's distribution)} \end{aligned}\right\}\ \text{Huxley and Zaazou (1949).}$$

If the values of W as a function of Z/p as found by Nielsen and Bradbury (1937) using an electrical shutter method be adopted, then it becomes possible to find the correct value of the constant factor in the above expression, and thus to obtain an indication of the prevailing law of distribution. Consequently it is possible, in principle, to determine whether the distribution changes from that of Maxwell, which must prevail as $Z/p \to 0$, to one more closely represented by Druyvesteyn's or some other distribution at the larger values of Z/p.

Comparison of the results of the two experiments does in fact show that good agreement may be obtained in hydrogen at the higher values of Z/p only if a law of distribution approximating to that of Druyvesteyn be assumed, but neither method of measuring W has been performed to a sufficient degree of accuracy at the lower values of Z/p to warrant any reliable conclusions being drawn as to the range of values of Z/p over which the transition occurs.

Table 3 shows the agreement between the measured values of W obtained from magnetic deflection and assuming Druyvesteyn's distribution and those obtained by Nielsen and Bradbury (N.B.).

Table 3

Z/p	$p=2$ mm	3 mm	4 mm	5 mm	6 mm	8 mm	N.B.
0·05						1·81 (1·79)	2·0
0·1				2·98 (2·95)	2·94 (2·98)	3·02 (3·01)	3·2
0·2			4·45 (4·53)	4·55 (4·45)	4·54 (4·48)		4·8
0·5	6·96 (7·02)	6·98 (6·95)	7·03 (6·98)	6·92 (7·02)			7·3
1·0	9·70 (9·71)	9·63 (9·68)	9·68 (9·65)				9·9
4·0	20·6						21·3

The values in brackets are for $B=32\cdot3$ gauss, and the others are for $B=21\cdot5$ gauss. The comparison is more easily made by reference to figure 4.

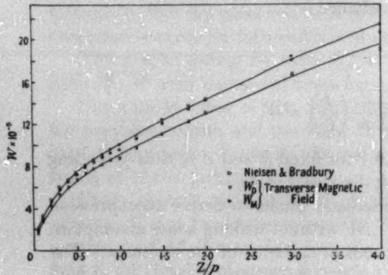


Figure 4. Drift velocities in hydrogen.

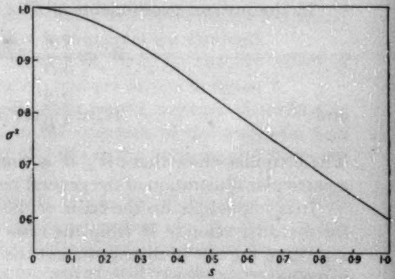


Figure 5. Graph of σ^2 as a function of s (assuming Maxwell's distribution).

Because the curve for the Druyvesteyn distribution almost coincides with that of Nielsen and Bradbury in the range of values of Z/p shown, it is not drawn. For $Z/p<0\cdot2$, the experimental error precludes the possibility of stating which is the actual distribution.

The field was obtained from a pair of Helmholtz coils. Other parameters remaining the same, ratios were always measured corresponding to each sense of the magnetic field. Since the ratio was found to be unaffected by the reversal of B it followed that there were no errors due to geometrical asymmetry of the anode or of the magnetic field.

§ 4. DIFFUSION AND DRIFT IN A LONGITUDINAL MAGNETIC FIELD

In this case the field B is applied in a direction parallel or antiparallel to that of the electric field Z.

Bailey (1930) investigated the effect of a longitudinal magnetic field on the electrons in the diffusion chamber and obtained measurements of the drift velocity W. In this work no account was taken of the distribution of the velocities of agitation of the electrons about the mean. The present experiment enables mean free paths to be measured taking into account the velocity distribution which is known from the experiments described in the previous section.

The apparatus used was that in which the receiving electrode is divided into annuli (figure 1 (*a*)). In the absence of a magnetic field the ratio R_b of the current falling on the annulus of radius b to the total current entering the chamber is given by (1). When a field B is applied parallel to Z, the divergence of the stream is reduced, consequently the ratio R_h is increased to a new value R_b'. Huxley (1940) has shown that R_b' is the ratio appropriate to an electrode with a central disc of radius b/σ with $B=0$, where $\sigma^2=K_B/K$. (K_B is the coefficient of diffusion normal to B.)

If the law of distribution of the speeds of agitation U is

$$f(U)\,dU=\frac{n}{\alpha^3\Gamma(3/n)}\exp\left\{\left(-\frac{U}{\alpha}\right)^n\right\}U^2\,dU$$

then it can be shown that

$$\sigma^2=1-\frac{\Gamma(2/n)}{\Gamma(4/n)}s^2+\frac{ns^4}{2\Gamma(4/n)}\int_0^\infty\frac{\exp(-t^{n/2})}{(t+s^2)}\,dt \qquad \ldots\ldots(5)\dagger$$

where $s=\omega l/\alpha$, and $\omega=Be/m$ where e and m are the electronic charge and mass respectively. This result follows from the formulae

$$K=\tfrac{1}{3}\overline{lU}\simeq\tfrac{1}{3}l\overline{U}$$

$$K_B=\frac{1}{3}\overline{\left[\frac{lU}{1+\omega^2T^2}\right]}\simeq\frac{l}{3}\overline{\left(\frac{U^3}{\omega^2l^2+U^2}\right)}$$

where it is assumed that the mean free path l has a constant value equal to its mean value appropriate to $f(U)$.

The case when $n=2$ corresponds to Maxwell's distribution and

$$\sigma^2=1-s^2-s^4\exp(s^2).\mathrm{Ei}(-s^2). \qquad \ldots\ldots(6)$$

The values of $\mathrm{Ei}(-s^2)$ were obtained from Jahnke and Emde's *Tables of Functions*, and the curve showing σ^2 as a function of s is given in figure 5. With $n=4$, corresponding to Druvyesteyn's distribution

$$\sigma^2=1-s^2\sqrt{\pi}+2s^4\int_0^\infty\frac{\exp(-x^2)}{x+s^2}\,dx \qquad \ldots\ldots(7)$$

and the value of the integral was calculated numerically, a similar curve being obtained for σ^2 as a function of s.

Now $s=\omega l/\alpha=1{\cdot}758\times10^7Bl/\alpha$, where B is measured in gauss. If Maxwell's law of distribution be assumed, then $\alpha=$ most probable velocity $=\overline{U}\sqrt{\pi}/2$ (Huxley and Zaazou 1949). Furthermore $[(\overline{U^2})]^{1/2}=1{\cdot}15\sqrt{k_1}\times10^7$ and

---

† I am indebted to Professor Huxley for this formula, which has not been published hitherto.

$\overline{U}/(\overline{U^2})^{1/2}=0{\cdot}92$, whence $\alpha=0{\cdot}94\times10^7\sqrt{k_1}$, and $s=1{\cdot}87Bl/\sqrt{k_1}$. On the other hand, according to Druyvesteyn's distribution

$$\overline{U^2}=\frac{4}{\alpha^3\Gamma(3/4)}\int_0^\infty \exp(-U/\alpha)^4\,.\,U^4\,dU$$

whence $\alpha^2=1{\cdot}35(\overline{U^2})$, and $(\overline{U^2})^{1/2}=1{\cdot}08\times10^7\sqrt{k_1}$; thus $s=1{\cdot}40Bl/k_1$. Thus for a given Z, p and B, to determine the value of the mean free path l, it is necessary first to measure the ratio R_b with $B=0$, thus obtaining Zh/k_1, and then to apply B and measure R_b'. σ can then be determined from the family of curves in figure 6.

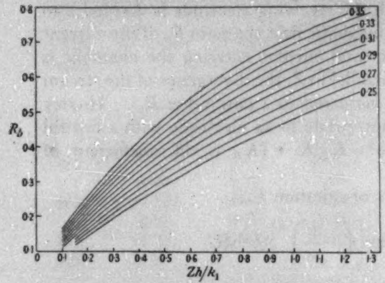


Figure 6. Ratio curves for varying values of b/h, for the electrode divided into annuli.

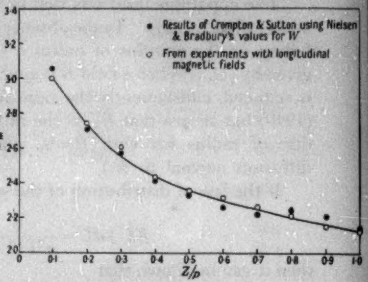


Figure 7. Mean free paths in hydrogen.

If we know σ, s can be found from figure 5, and hence l can be calculated. The mean free path at unit pressure $L=lp$, and good agreement is found between values of L for constant Z/p obtained for varying pressures and magnetic fields (table 4). The field B was provided by a pair of Helmholtz coils as in § 3. Results of greater accuracy could be obtained by using much stronger fields than those employed in the present case.

Table 4

$Z/p=0{\cdot}5$	$p=2$ mm	$p=4$ mm
$B=98{\cdot}9$ gauss	$L=2{\cdot}43\times10^{-2}$ cm	$L=2{\cdot}45\times10^{-2}$ cm
86·0	2·45	2·42
64·5	2·49	2·44

These values of L are calculated on the assumption of Druyvesteyn's law of velocity distribution, and may be compared with a value of $2{\cdot}44\times10^{-2}$ cm which is obtained for the mean free path at unit pressure L, for $Z/p=0{\cdot}5$, assuming Crompton and Sutton's values for k_1 and Nielsen and Bradbury's values for W. From figure 7 it can be seen that the values of the mean free path at unit pressure show excellent agreement with the values calculated from the formulae

$$L=7{\cdot}11\times10^{-9}\frac{W(k_1)^{1/2}}{Z/p}\text{ cm}\quad\text{(Maxwell)}$$

$$L=7{\cdot}38\times10^{-9}\frac{W(k_1)^{1/2}}{Z/p}\text{ cm}\quad\text{(Druyvesteyn)}$$

using values of k_1 obtained by Crompton and Sutton and Nielsen and Bradbury's values for W.

ACKNOWLEDGMENTS

The author wishes to express her thanks to Professor Huxley who suggested and encouraged these researches, and to Dr. D. J. Sutton for his helpful guidance throughout. She also wishes to thank the University of Adelaide for a research scholarship, during the tenure of which this work was carried out. The author is at present the holder of a C.S.I.R.O. studentship, which is permitting her to continue work in this field.

REFERENCES

BAILEY, V. A., 1930, *Phil. Mag.*, **9**, 560, 625.

CROMPTON, R. W., HUXLEY, L. G. H., and SUTTON, D. J., 1953, *Proc. Roy. Soc.* A, **218**, 507.

CROMPTON, R. W., and SUTTON, D. J., 1952, *Proc. Roy. Soc.* A, **215**, 467.

HUXLEY, L. G. H., 1937, *Phil. Mag.*, **23**, 210, 422; 1940, *Ibid.*, **29**, 313; 1951, *Proc. Phys. Soc.* B, **64**, 844.

HUXLEY, L. G. H., and CROMPTON, R. W., 1955, *Proc. Phys. Soc.*, B, **68**, 381.

HUXLEY. L. G. H., and ZAAZOU, A. A., 1949, *Proc. Roy. Soc.* A, **196**, 402.

NIELSEN, R. A., and BRADBURY, N., 1937, *Phys. Rev.*, **51**, 69.

APPENDIX

BY L. G. H. HUXLEY

The following supplementary comments on the subject matter of Miss Hall's paper may be of interest.

I. *The equivalent collisional cross section of the molecules of a gas for the processes of diffusion and drift*

Measurements of the spread of a stream of electrons diffusing in a gas and drifting in a uniform electric field Z, permit the ratio W/K of the drift velocity to the coefficient of diffusion to be directly determined as some function $F(p, Z/p)$ of the pressure p of the gas (at 15°c) and the ratio Z/p. If W is separately measured as a function of Z/p then it follows that K may be separately expressed as a function of p and Z/p. It has been found convenient to interpret the macroscopic experimental quantities W and K in terms of the microscopic parameters of a simplified model gas in which the W and K in the model are the same functions of p and Z/p as those measured in the actual gas. The model that has been used most frequently is the following. The trajectory of an electron through the gas comprises a sequence of rectilinear segments with abrupt changes in direction at the junction of successive segments. Thus, the trajectory is that which an electron, regarded as a small particle, would follow were the molecules to behave as rigid particles, but this interpretation is of no theoretical consequence since the model is essentially geometrical and not dynamical. It is moreover assumed for simplicity that all directions of motion following the terminating deflections at the ends of a set of free paths, all parallel to a fixed direction, are equally probable. That is to say the electrons considered as particles are scattered isotropically in collisions with molecules as if the latter were small, smooth, rigid and, relative to the electrons, massive spheres. If l_0 is the mean free path and N the number of scattering centres in unit volume then the equivalent scattering cross section of a molecule is

$$A = 1/Nl_0.$$

The application of the theory of diffusion and drift to this model leads to the following well-known formulae for the diffusion coefficient K and drift velocity W of the electrons

$$\left.\begin{aligned} K &= \tfrac{1}{3}\overline{l_0 U} \quad \dots (a) \\ W &= \tfrac{2}{3}\frac{Ze}{m}\overline{\left(\frac{l_0}{U}\right)} \dots (b) \end{aligned}\right\} \qquad \dots\dots(\mathrm{A1})$$

where U is the agitational speed of an electron.

In general the assumption of isotropic scattering will not correspond to the physical facts and formulae (A1) require further discussion.

Many years ago Townsend (*Electricity in Gases*, § 83) discussed non-isotropic scattering in relation to the diffusion of massive ions in a gas, but the argument, being geometrical in character may be applied to non-isotropic scattering in general. It is found that formulae (A1) are still correct (except that Townsend's formula for W does not contain the factor 2/3) if l_0 is replaced by

$$l = l_0 + l_1. \qquad \dots\dots(\mathrm{A2})$$

The correction l_1 to the free path l_0 when the molecular concentration N is specified would demand detailed calculation in each gas studied, using the dynamical theory of gases or the quantum mechanics of collisions and it would be implicit in any formula for K so derived. Such, however is a separate study and is not our present concern. The equivalent cross section for scattering now becomes $A = 1/Nl$ and the case of non-isotropic scattering is thus expressed in terms of a model based on the assumption of isotropic scattering by replacing l_0 by l. The quantity l_0 for the model gas has the following interpretation: Consider a large group of electrons starting normal to a plane and moving into the model gas. After each electron has received a collision the members of the group move isotropically and its centroid is brought to rest at a distance l_0 from the plane. However, when the group is projected into an actual gas, although the specification of what is or is not a collision may be difficult or imprecise, yet the members of the group will ultimately move isotropically and the centroid will be brought to rest at some distance l from the starting plane. This is the distance $l = l_0 + l_1$ that is to be used in formulae (A1).

It is evident from the geometrical basis of the model upon which formulae (A1) are based that nothing is postulated concerning exchanges of energy between electrons and gas molecules beyond the fact that in a steady state of motion in a uniform electric field Z, electrons must lose energy to the gas at the mean rate ZeW per electron and that neither the magnitude of the mean energy of agitation $\tfrac{1}{2}m\overline{U^2}$ of the electrons nor the law of distribution of the speeds U can be deduced for it.

In order therefore to deduce the value of l from equations (A1) it is necessary to make use of the experimental measurements of W and $2\lambda = W/K$.

It is assumed for convenience in considering the mean values $\overline{lU}$ and $(\overline{l/U})$ that l does not change rapidly with U and that the law of distribution of the speeds U is sufficiently narrow to justify the approximations $l\overline{U}$ and $l(\overline{U^{-1}})$ respectively to these averages.

With these approximations it follows that

$$\left.\begin{aligned} & & 2\lambda &= W/K = (Ze/\tfrac{1}{2}m\overline{U^2})f \quad \ldots(a)\\ &\text{where} & f &= (\overline{U^2}\,\overline{U^{-1}})/\overline{U} \quad \ldots(b)\\ &\text{also (A1}a) & W &= \frac{2}{3}\frac{Ze}{m}l(\overline{U^{-1}}) \quad \ldots(c). \end{aligned}\right\} \quad \ldots\ldots(\text{A3})$$

Equation (A3a) is exact in the limiting case of $W \to 0$, and as it is independent of the properties of the model expressed through l, it is probably exact or almost so for any gas in general. Since there are three unknown quantities, the mean kinetic energy $\frac{1}{2}m\overline{U^2}$, the dimensionless factor f whose value is determined by the law of distribution of the speeds U, and the mean free path l, but only two experimental quantities 2λ and W from which to determine them, it is clear that some further postulate or some additional experimental quantity must be measured if these three unknowns are to be determined. The simplest procedure is to postulate alternative and probable laws of distribution, for instance that of Maxwell or of Druyvesteyn and to deduce the values of $\frac{1}{2}m\overline{U^2}$ and $l = 1/NA$ on the basis of these postulates. It is found, for instance that A, considered as a function of $\overline{U}$, is relatively insensitive to the law of distribution adopted.

Another and more satisfactory procedure is to measure a further physical quantity associated with the motion of the electrons in the gas such that its behaviour can be expressed by means of a formula based on precisely the same model as that employed above. If a further physical quantity can be measured then the self-consistency of the model may be checked with respect to the three parameters, $\frac{1}{2}m\overline{U^2}$, A and f. Miss Hall's measurements of the magnetic deflection of an electron stream and of the diffusion in a longitudinal magnetic field provide the measurements of the additional quantities required for those purposes.

It may be remarked that it would be illogical to suggest that formulae (A1) and (A3) should be adjusted to conform with other formulae based on other models or derived on other principles, for as explained above, formulae (A1) and (A3) serve to define the equivalent cross section of the gas molecules and when applied to the experiment values of W and K yield the equivalent cross section of the molecules for collision with electrons and the electron temperature as functions of Z/p.

The model is moreover a useful transfer device for making available the macroscopic properties of electronic motion in gases in situations where direct experiments are not possible. Thus the formulae based on this model may be used to deduce the macroscopic behaviour of electrons in the ionosphere under the action of the alternating electric field of an electromagnetic wave, as for instance in the theory of radio wave interaction.

II. *Comment on the differential equation for the electron concentration in the presence of a magnetic field*

The equation of continuity for the concentration n of the electrons is

$$\frac{\partial n}{\partial t} + \operatorname{div} n(\mathbf{W} + \mathbf{w}) = 0 \qquad \ldots\ldots(\text{A4})$$

where $n\mathbf{W}$ is the flux density of electrons due to drift in an electric field and $n\mathbf{w}$ that due to diffusion. It is customary to assume that the cartesian components of $n\mathbf{w}$ are linear functions of the components of grad n. Also when $\mathbf{W}$ is zero a group of electrons with axial symmetry about an axis parallel to the magnetic field $\mathbf{B}$ retains that symmetry as it spreads by diffusion. Thus, in the presence of a magnetic field $\mathbf{B}$ parallel to Oz, the vector $n\mathbf{w}$ may be written

$$-n\begin{Bmatrix} w_x \\ w_y \\ w_z \end{Bmatrix} = \left\|\begin{matrix} K_B & -K_B' & 0 \\ K_B' & K_B & 0 \\ 0 & 0 & K \end{matrix}\right\| \begin{Bmatrix} \dfrac{\partial n}{\partial x} \\ \dfrac{\partial n}{\partial y} \\ \dfrac{\partial n}{\partial z} \end{Bmatrix} \qquad \ldots\ldots(\mathrm{A5})$$

where K_B, K_B' and K are coefficients of diffusion.
When $\mathbf{B}=0$, then $K_B=K$; $K_B'=0$.

It follows from equations (A4) and (A5) that when $\mathbf{W}$ is independent of position and of time, then

$$K_B\left(\frac{\partial^2 n}{\partial x^2}+\frac{\partial^2 n}{\partial y^2}\right)+K\frac{\partial^2 n}{\partial z^2}=\mathbf{W}.\operatorname{grad} n+\frac{\partial n}{\partial t}. \qquad \ldots\ldots(\mathrm{A6})$$

The differential equation for a steady state distribution of n follows from (A6) by equating $\partial n/\partial t$ to zero.

It may be remarked that the coefficient K_B' does not appear in (A6) nor does it contain any explicit term corresponding to a supposed volume force density $e(n\mathbf{w}\times\mathbf{B})$.

The Influence of Temperature on Magnetic Viscosity

By J. H. PHILLIPS†, J. C. WOOLLEY† and R. STREET‡
†The University, Nottingham
‡ The University, Sheffield

Communicated by L. F. Bates ;
MS. received 28th July 1954 and in amended form 28th December 1954

Abstract. Measurements have been made to demonstrate the effect of temperature on irreversible magnetic viscosity in different types of precipitation alloys: Pt–Co, Ni_3Au and Alnico in an undeveloped state. In this type of alloy, the intensity of magnetization varies as a function of time according to the law $I=S_0 \ln t$ + const. On the activation energy model of irreversible magnetic viscosity if ΔE is the change in activation energy produced by a small change in applied field ΔH, $\Delta E=q\Delta H$. For the Alnico and Ni_3Au specimens the two magnetic viscosity parameters S_0 and $\boldsymbol{k}T/q$ are approximately linearly related to temperature. For Pt–Co the viscosity parameters and the coercivity decrease rapidly with increasing temperature. The results are interpreted in terms of Néel's disperse field theory of magnetization with the additional assumption that strains set up between coherent precipitate and matrix are responsible for the production of the disperse fields. It is predicted from the observations of magnetic viscosity that in Pt–Co the disregistry between precipitate and matrix decreases on increasing the temperature. This has been verified by measurements of the differential thermal expansion between the material of precipitate and the matrix by x-ray methods.

§ 1. Introduction

The present communication is an extension of the work previously reported by Phillips, Street and Woolley (1954, to be referred to as I). The relevant background will therefore be given only briefly here and reference should be made to I for more detailed consideration. Experimentally the effect of irreversible magnetic viscosity is such that in a constant applied field, almost invariably the intensity of magnetization I varies as a function of time t according to the law

$$I=S_0 \ln t + \text{const.} \qquad \ldots\ldots(1)$$

where t is measured from the instant the applied field becomes fixed and S_0 is a constant characteristic of the state of the material, its temperature and the point of measurement on the magnetization curve.† It is generally agreed that the underlying physical mechanism responsible for irreversible magnetic viscosity phenomena is the effect of thermal agitation on the processes of magnetization. Similar theoretical analyses have been given by Street and Woolley (1949) and Néel (1950, 1951).

---

† In practice it is usually necessary to correct for the effects of demagnetizing fields, as discussed by Street, Woolley and Smith (1952). The results presented here have been corrected in this way.

In the first of these, the effect of thermal agitation as an activation process is considered. Each irreversible magnetization process requires a certain activation energy E. By assuming that over limited ranges of E the number of such processes with activation energies in the range E to $E+dE$ is a constant p, it may be shown that

$$S_0 = \bar{\imath} p \boldsymbol{k} T \qquad \ldots\ldots(2)$$

where $\bar{\imath}$ is the mean contribution to the intensity of magnetization of the specimen resulting from a successful activation and T is the absolute temperature.

The effect of small changes ΔH in the applied field can be represented by a change in activation energy ΔE and it is found experimentally that

$$\Delta E = q \Delta H. \qquad \ldots\ldots(3)$$

q and S_0 are related by the equation

$$q = \chi_{\mathrm{irr}} \boldsymbol{k} T / S_0 \qquad \ldots\ldots(4)$$

where χ_{irr} is the irreversible susceptibility at the point of measurement. In Néel's analysis, the variation of I as a function of time is shown to be

$$I = \chi_{\mathrm{irr}} S_{\mathrm{v}} \ln t + \text{const.} \qquad \ldots\ldots(5)$$

Hence from equations (1), (4) and (5) the constant S_{v} in Néel's analysis is related to q by the relation

$$S_{\mathrm{v}} = \boldsymbol{k} T / q. \qquad \ldots\ldots(6)$$

The form of equation (3) relating energy changes to field changes suggests that q may be interpreted in terms of the volume of material involved in certain magnetization processes. By considering the energy variation of 180° boundary wall movements it can be shown that

$$q = 2kJ_s \Delta\bar{v} \qquad \ldots\ldots(7)$$

where k depends on the orientation of the boundary wall with respect to the field, and has a maximum value of 1, J_s is the spontaneous magnetization and $\Delta\bar{v}$ is the mean volume swept out by the boundary wall before irreversible displacement occurs.

The influence of temperature on magnetic viscosity is obviously of importance although relatively few investigations on a number of different materials have been carried out. In I it was demonstrated that, for Alnico and Fe_2NiAl, S_0 is an accurately linear function of the absolute temperature of measurement in the range 90°K to 900°K. Thus for these materials the product $\bar{\imath}p$ of equation (2) is independent of temperature. For Alnico q is constant, within the limits of experimental error, at temperatures above 300°K, but it is probable that from 90°K to 300°K the results are in better agreement with a law of the form $q \propto T^{1/4}$. The measurements described in I were made at the coercive force point of an Alnico of coercivity 356 oersteds. Barbier (1953) investigated the influence of temperature on S_{v} (equation (6)) using an Alnico specimen of almost the same coercivity. His measurements were made in the Rayleigh region of magnetization over the temperature range 20°K to 373°K and again could be represented by a law of the form $S_{\mathrm{v}} \propto T^{3/4}$, i.e. $q \propto T^{1/4}$. Courvoisier (1945) investigated a steel specimen over the range 193°K to 368°K and his results show that q is independent of temperature. However the last result is of little significance since the temperature range is not sufficiently large.

The work described below was undertaken to extend measurements on the influence of temperature on magnetic viscosity to materials other than the previously studied Alnico of coercivity 350 Oe approximately.

§ 2. RELEVANT PROPERTIES OF ALLOYS INVESTIGATED

The specimens examined in this investigation were alloys of Pt–Co, Ni–Au and Alnico which had been heat treated to have relatively low coercivity. All these materials are precipitation alloys, i.e. at high temperatures the equilibrium state is single phase but on cooling to lower temperatures separation into two or more phases occurs. The extent of precipitation may be controlled by the speed of cooling from the high temperature state and also by annealing at temperatures below the solution temperature. Alnico is a complex alloy and the nature of the precipitate has not been determined precisely. Metallurgically the binary Ni–Au system is much simpler as there is a single broad miscibility gap extending from about 10–98 atomic % Ni, the maximum occurring at 840°c for 70 atomic % Ni (Bozorth 1951). Above the gap there is a continuous series of solid solutions having face-centred cubic structure. In the miscibility gap the equilibrium state is two phase consisting of nickel rich and gold rich face-centred cubic phases. The special interest of Ni–Au alloys here is that it is possible to produce precipitation alloys in this system for which the Curie temperature is lower than the solution temperature. Thus it has been possible to study magnetic viscosity at temperatures close to the Curie temperature without introducing metallurgical changes during measurement.

The Pt–Co system has been studied by Gebhardt and Köster (1940) and by Newkirk *et al.* (1950, 1951). At lower temperatures in the range 25–55 atomic % Co the alloys in equilibrium are ordered and have face-centred tetragonal structure. The boundary of the ordered phase region has a maximum at 825°c for the 50–50 atomic % alloy. At higher temperatures the equilibrium state is single phase with face-centred cubic structure and this may be retained at lower temperatures in non-equilibrium conditions by sufficiently rapid quenching.

The precipitation alloys described above may be developed by suitable heat treatment to have relatively high coercivities, i.e. for Alnico ~500 Oe, for 50–50 % Pt–Co ~3000 Oe and for Ni_3Au ~150 Oe. Geisler (1953) has given an empirical relation between coercivity and the disregistry between the precipitate and the surrounding matrix. The disregistry is measured as the percentage difference of the lattice parameters of the matrix and precipitate if these have the same structure. If the two crystal structures are different then disregistry may be measured as the percentage difference of the areas of the conjugate planes of the precipitate and matrix. Geisler (1949, 1953) supposed that in the earlier stages of precipitation for which the coercivity has the greatest values, the precipitate and matrix are coherent and can give rise to large internal strains. In I these highly localized internal strains were considered to produce fluctuations in the directions of the elementary spontaneous magnetization vectors setting up regions of disperse field as considered by Néel (1946) in his theory of coercive force. A region of disperse field effectively acts as a magnetic unipole, possessing a high magnetostatic energy, when not intersected by a domain boundary wall, but when it is intersected it is converted into a dipole of lower magnetostatic energy. Thus in the course of magnetization to move a boundary wall away from an intersection with a region of disperse field energy must be provided by the external field to convert the dipole into an effective unipole. In the absence of detailed information of the metallurgical processes occurring during precipitation it is not possible to develop these suggestions more precisely. However it has been shown that the results of the measurements described in I, on the progressive development of precipitate in precipitation alloys, are in general agreement with the above summary. Particular consideration of the interpretation of q in equation (7) shows that $\Delta\bar{v}$ is related to the mean volume of the regions of disperse field. Thus, for highly localized disperse field regions set up by coherent precipitates, $\Delta\bar{v}$ and hence q will have relatively low values. Development of precipitation by heat treatment eventually leads to the breakdown of coherency. Consequently the highly localized intense coherency strains disappear leaving less intense non-localized disperse field regions, set up for example by quenching strains, which may spread over larger volumes. Thus development of the precipitate above a critical size results in decreasing coercivity and increasing values of q.

§ 3. METHOD OF MEASUREMENT

Measurements of the time variation of the intensity of magnetization of the rod specimens were made as described in I, by a magnetometric method using a fluxgate. The maximum magnetic field available from the solenoid used for magnetizing the specimens was 1300 Oe, and for this reason it was necessary to limit measurements to specimens having coercivities of a few hundred oersteds.

S_0 and χ_{irr} were determined as functions of temperature from which $\boldsymbol{k}T/q = S_0/\chi_{\text{irr}}$ from equation (4) was calculated.

The specimens were heated in a non-inductively wound electrical furnace mounted inside the magnetizing solenoid. During the measurements it was necessary to maintain the specimens at elevated temperatures for appreciable periods of time. To ensure that metallurgical changes did not occur during such time the room temperature values of the magnetic parameters were checked after each high temperature run. If these values differed significantly from the initial room temperature values, measurements were discontinued.

§ 4. RESULTS

4.1. *Platinum Cobalt*

The alloy was of 50–50 atomic percentage composition in the form of rods 19 cm long and 0·39 cm in diameter. Preliminary measurements showed that the magnetic parameters varied with heat treatment in a way similar to that of the Alnico type of alloy described in I. From these observations it was possible to determine the heat treatment required to produce specimens having relatively low coercivities of the order of a few hundred oersteds at room temperatures.

Typical values of S_0, $\boldsymbol{k}T/q$ and H_c as functions of absolute temperature are plotted in figure 1 from which it is seen that all three quantities decrease rapidly with increasing temperature. This is in direct contrast to the behaviour of the Alnico and similar specimens described in I for which S_0 and $\boldsymbol{k}T/q$ increase almost linearly with temperature while H_c decreases approximately in the same way as the spontaneous magnetization.

Measurements were also made on 50–50 atomic % Pt–Co specimens subjected to heat treatments producing coercivities of 100 Oe and 600 Oe at room temperature. In both cases the temperature variations of S_0, $\boldsymbol{k}T/q$ and H_c were of the same form as those in figure 1.

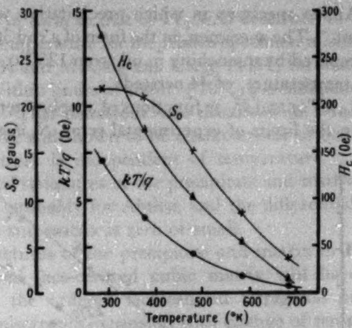


Figure 1. Variation of S_0, $\boldsymbol{k}T/q$ and H_c with temperature for Pt–Co.

4.2. *Nickel Gold*

The alloy, of 25 atomic % Au, 75 atomic % Ni, corresponding to Ni_3Au, was in the form of rods 22 cm long and 0·6 cm in diameter. For this composition the solid solution temperature is 1113°K and the Curie temperature 613°K. Preliminary investigation of the effect of heat treatment on magnetic properties gave similar results to those described in §4.1. However the values of S_0 were much smaller than those for Pt–Co and Alnico; but this is to be expected as the spontaneous magnetization of the Ni_3Au alloy is relatively low.

The values of S_0, $\boldsymbol{k}T/q$ and H_c as functions of absolute temperature are shown in figure 2 for a Ni_3Au specimen having a coercivity of 110 Oe at room temperature.

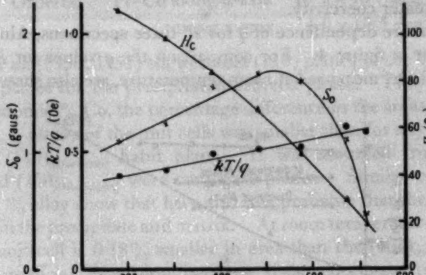


Figure 2. Variation of S_0, $\boldsymbol{k}T/q$ and H_c with temperature for Ni_3Au.

From room temperature to about 450°K, S_0 increases approximately linearly with temperature according to a law of the form $S_0 = \text{const.} \times T$, but above 450°K S_0 decreases rapidly. Again $\boldsymbol{k}T/q$ is almost a linear function of T in the range 290°K to 560°K, and H_c decreases relatively slowly with increasing temperature.

4.3. *Alnico*

Previous investigations of the temperature dependence of magnetic viscosity in Alnico were all on specimens of relatively high coercivity corresponding to an appreciable development of the precipitate phase. The present results were obtained with an Alnico specimen in which precipitation was in a much earlier stage of development. The specimen, in the form of a rod 30 cm long and 0·6 cm in diameter, was prepared by quenching in oil from 1250°c, and had a coercivity, measured at room temperature, of 44 oersteds.

The values of S_0, $\boldsymbol{k}T/q$ and H_c as functions of absolute temperature are plotted in figure 3. Within the limits of experimental error, S_0 is a linear function of T and over the limited range 290°K to 510°K, $\boldsymbol{k}T/q$ is also linear. H_c increases with increasing temperature in direct contrast to the behaviour observed for Alnico specimens of greater coercivity.

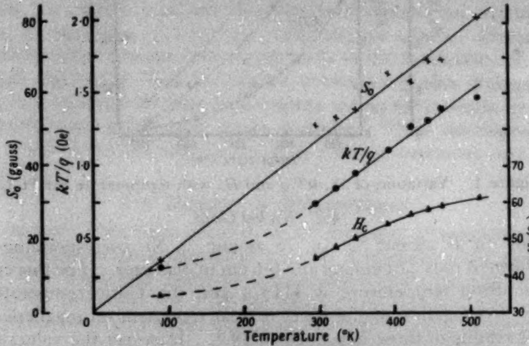


Figure 3. Variation of S_0, $\boldsymbol{k}T/q$ and H_c with temperature for an undeveloped specimen of Alnico.

The temperature dependence of q for all three specimens, Alnico, Ni_3Au and Pt–Co, is shown in figure 4. For comparison the q values for Alnico having a coercivity of 350 Oe, measured at room temperature, are also shown in this figure.

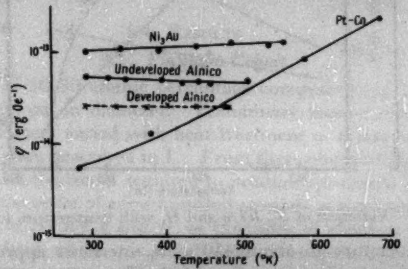


Figure 4. Variation of q as a function of absolute temperature for Ni_3Au, Pt–Co and undeveloped Alnico. The results for Alnico of coercivity 350 Oe at room temperature from measurements by Street, Woolley and Smith (1952) are included for comparison.

§ 5. Discussion

The results given in §4 for Alnico and Ni_3Au are generally similar to those already described in I in that both S_0 and $\boldsymbol{k}T/q$ are approximately linear functions of absolute temperature. The relatively slow variation of H_c with temperature for Ni_3Au is similar in form to that of Fe_2NiAl described in I. The anomalous increase in H_c with temperature for the Alnico specimen is similar to that reported by Jellinghaus (1943) and McCaig and Clegg (1954) for materials in the under-developed state. Within limits of about $\pm 10\%$, q is independent of temperature for both materials. This behaviour was discussed in I in terms of the ideas described in § 2 above and shown to be characteristic of materials for which the state of coherent strain is independent of temperature. This condition will hold when the crystal structures of the precipitate and matrix are cubic, as is the case for Ni_3Au and probably for Alnico, and the differential thermal expansion between precipitate and matrix is zero or small.

The crystal structures of the precipitate and matrix in Pt–Co are dissimilar (§ 2); the disordered face-centred cubic matrix will have isotropic thermal expansion whereas the ordered face-centred tetragonal precipitate might be expected to show anisotropic expansion with change of temperature. Thus the state of coherency strain (see Geisler 1949, 1953) in this alloy between precipitate and matrix will inevitably be temperature dependent. The very rapid temperature variation of q (increase) and H_c (decrease) strongly suggests that the coherency strains decrease as the temperature increases (§ 2). In order to verify this prediction x-ray analysis was carried out on small wire samples, taken from the Pt–Co rods used in the magnetic investigation, to determine the expansion coefficients of the ordered and disordered phases. Measurements were made over the temperature range 20°c to 400°c in an evacuated high temperature x-ray powder camera, and the approximate linear expansion coefficients were determined as

Disordered Pt–Co	$10{\cdot}9 \times 10^{-6}$ deg^{-1}
Ordered Pt–Co along *a*-axis	$10{\cdot}9 \times 10^{-6}$ deg^{-1}
along *c*-axis	$4{\cdot}6 \times 10^{-6}$ deg^{-1}

Newkirk *et al.* (1950, 1951) showed that (110) planes of the disordered matrix are the habit planes for the precipitated ordered phase. They showed that for an alloy of 54 atomic % Co, the percentage difference in the areas of the $(101)_{\text{ordered}}$ and $(110)_{\text{disordered}}$ planes of the unit cells was smaller than for any other pair which included the $(110)_{\text{disordered}}$ habit plane. It was assumed from this that the $(101)_{\text{ordered}}$ and $(110)_{\text{disordered}}$ were conjugate planes. Similar considerations for the 50 atomic % alloy show that here also it is probable that the same two planes are conjugate in the precipitate and matrix. At room temperature the $(110)_{\text{disordered}}$ plane of the unit cell is $0{\cdot}18\%$ smaller in area than the $(101)_{\text{ordered}}$ plane. The differential superficial expansion between these conjugate planes, calculated from the coefficients of linear expansion given above show that this difference in area will be reduced as the temperature increases. Calculation shows that the areas of the conjugate planes are equal at about 600°c. Thus the x-ray results are in direct agreement with the magnetic observations showing that the volume of Pt–Co affected by disperse field regions (of decreasing strength) increases as temperature increases.

For Ni_3Au and undeveloped Alnico there are much smaller regular variations of q as functions of temperature as may be seen from figure 4. It is possible that these variations also are due to differential thermal expansion between conjugate planes of the precipitate and matrix in these alloys. However it is unlikely that x-ray measurements would be sufficiently precise to demonstrate the differential expansion directly.

The above discussion has been specifically confined to a consideration of the disperse fields set up by strain. Néel (1946) has shown that disperse fields may also be set up independently of strain effects, by inclusions or cavities in which the spontaneous magnetization is different from that of the surrounding matrix. This type of mechanism will also be operative to some degree in the materials studied here, but it appears difficult to account for the experimental results unless it is assumed that the strain mechanism is of primary importance. Additional support for the assumption is given by the work of Geisler (1953) who showed that the coherency strains due to disregistry are more significant in producing high coercivity precipitation alloys, than, for example, the differences in the spontaneous magnetization of the precipitate and matrix.

ACKNOWLEDGMENTS

The authors wish to thank Professor L. F. Bates for his interest and continued support of this work, and for the use of the high temperature x-ray camera purchased by him with a Royal Society grant. Their thanks are also due to Imperial Chemical Industries Ltd. for the loan of electronic equipment and to Dr. Alun Edwards and the Permanent Magnet Association for provision of the Alnico specimens. The Pt–Co and Ni–Au specimens were provided on loan by the Sheffield Smelting Company Ltd., and thanks are due to Mr. R. C. Jewell and his staff for their cooperation, and for successfully overcoming the many difficulties involved in manufacturing the rods of these alloys. One of us (R. S.) thanks the Royal Society for a grant for the purchase of a recording potentiometer.

REFERENCES

BARBIER, J. C., 1953, *Thèse*, L'Université de Grenoble.

BOZORTH, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand), p. 314.

COURVOISIER, P., 1945, *S.B. Bayer. Akad. Wiss.*, **10,** 89.

GEBHARDT, E., and KÖSTER, W., 1940, *Z. Metallk.*, **32,** 253.

GEISLER, A. H., 1949, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **180,** 230; 1953, *Rev. Mod. Phys.*, **25,** 316.

JELLINGHAUS, W., 1943, *Arch. Eisenhuttenw.*, **16,** 247.

MCCAIG, M., and CLEGG, A. G., 1954, *BEAIRA Report*, N/T67.

NÉEL, L., 1946, *Ann. Univ. Grenoble*, **22,** 299 ; 1950, *J. Phys. Radium*, **11,** 49 ; 1951, *Ibid.*, **12,** 339.

NEWKIRK, J. B., SMOLUCHOWSKI, R., GEISLER, A. H., and MARTIN, D. L., 1950, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **188,** 1249 ; 1951, *J. Appl. Phys.*, **22,** 290.

PHILLIPS, J. H., STREET, R., and WOOLLEY, J. C., 1954, *Phil. Mag.*, **45,** 505.

STREET, R., and WOOLLEY, J. C., 1949, *Proc. Phys. Soc.* A, **62,** 562.

STREET, R., WOOLLEY, J. C., and SMITH, P. B., 1952, *Proc. Phys. Soc.* B, **65,** 679.

Surface Recombination in Germanium in the Presence of Strong Electric Fields

By H. K. HENISCH and W. N. REYNOLDS
Department of Physics, University of Reading

Communicated by R. W. Ditchburn; MS. received 22nd November 1954

Abstract. Strong electric fields have been applied in a direction normal to the surfaces of germanium single crystals; such fields have been shown to have a large effect on the surface recombination velocity of injected minority carriers. The variation is in the same sense for n-type and p-type samples. A possible theoretical model for this asymmetry is discussed.

---

§ 1. INTRODUCTION

MINORITY carriers injected into a semiconducting filament decay either in the bulk of the material or at the surface. If the bulk lifetime of the carriers is known and not too small, measurements of the effective lifetime of minority carriers in a filament can be used to evaluate the rate of decay at the surface. This is expressed as the surface recombination velocity s, as defined by Shockley (1950). To obtain more detailed information on the processes involved, it is necessary to alter the surface conditions so as to produce systematic changes of s. Many experiments, for example those of Brattain (1951), show that a potential barrier exists at a free etched surface which can be ascribed to the presence of charged surface states. One may then expect the surface recombination process to be influenced by the parameters of this barrier or, at any rate, by the distribution and occupancy of the surface states. External influences designed to vary these quantities systematically should lead to corresponding changes of s. Such changes can be produced by varying the ambient gas atmosphere, for example as shown by Keyes and Maple (1954). The present experiments were carried out to provide more information on these problems. Instead of varying the ambient gas atmosphere, a strong electric field was applied between the surface under test and an external metal plate. The surface recombination velocity was then determined as a function of this field. This procedure is completely reversible and does not involve permanent damage to the surface under investigation.

§ 2. EXPERIMENTAL METHOD

The measurements of filament lifetime τ_f were performed using the pulse method described by Many (1954). In this method, the minority carriers are injected through soldered end contacts. The single crystal specimen is connected as one arm of a bridge and its changing resistance with time during the pulse is balanced by an *RC* combination in another arm. The specimens had resistivities between 5 and 12 ohm cm and were prepared in the form of long thin plates of about 0·05 cm × 0·4 cm cross section and 1·5 to 3 cm length. The surface recombination velocity can then be calculated from the bulk lifetime and the dimensions as shown by Shockley (1950). The bulk lifetimes, about 75 μ sec for n-type and 40μ sec for p-type material, were determined by previous measurements on much thicker specimens. The external electric field was applied as shown in figure 1. A thin leaf of mica ($\sim$15–20μ) was used as an insulating spacer. The potential drop along the filament was negligible compared with the voltage applied to the external plates. All experiments were performed in darkness in air dried by P_2O_5. The experiments under discussion were carried out on surfaces etched with C.P.4 (Haynes and Shockley 1951). Thick oxide layers of the type produced by repeated cycling of the ambient gas atmosphere (Brattain and Bardeen 1953) are not to be expected here.

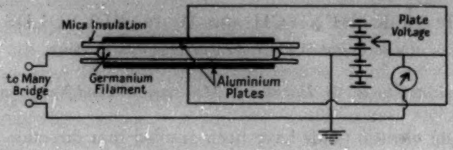


Figure 1. Experimental circuit.

§ 3. Experimental Results

Figures 2 and 3 show the general behaviour of etched n-type and p-type surfaces in the presence of strong electric fields at several different temperatures. The evaluation of these results requires a knowledge of the bulk lifetime as a function of temperature. This was determined by separate experiments on large well-etched specimens, cut from positions in the crystal adjacent to those of the filaments used in the principal experiment described above. As far as could be judged without accurate timing arrangements, the changes of *s* are complete within about 0·1 sec after the application of the plate voltage. Specimens which are ground or sand-blasted show no change of filament lifetime with applied field. The magnitude of the external voltage applied was limited by the dielectric breakdown strength of the mica used. The experiments confirm a previous result (Reynolds 1953) according to which the surface recombination velocity for zero field increases sharply as the temperature falls below room temperature.

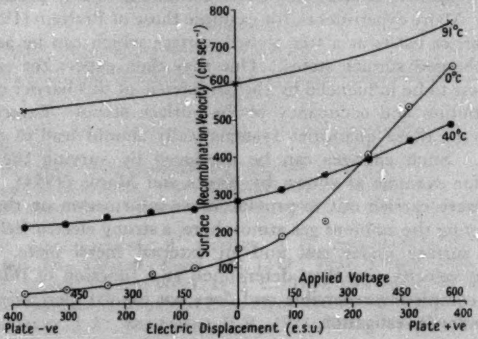


Figure 2. Experimental results for n-type germanium.

The total number of additional carriers and hence the number of carriers recombining is much greater at the injecting end of the filament than at the other end. The region near the injecting end should therefore be much more sensitive to the applied field than the remainder of the specimen. This has been confirmed by a subsidiary experiment, using a metal plate which was split into two sections insulated from one another, one near to the injecting end contact and one further away from it. Similar methods can also be used for a transistor demonstration in which the emitter current is kept constant and in which the transmission probability of injected carriers to the collector is modulated by the external field (Henisch, Reynolds and Tipple 1955).

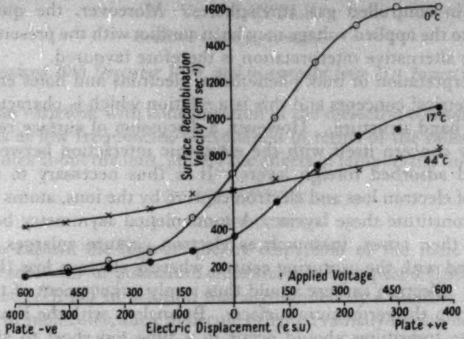


Figure 3. Experimental results for p-type germanium.

§ 4. DISCUSSION

The mechanism of surface recombination is not yet fully understood, but the following considerations may form the basis of a theoretical interpretation.

The principal feature of the results shown in figures 2 and 3 is the fact that the observed changes of *s* are in the same direction for n-type and p-type material at all temperatures. Rectification tests carried out with unformed point contacts on these particular surfaces showed pronounced and almost identical rectification properties, subject to the usual change of polarity. This suggests that symmetrical contact barriers exist on the two types of material under these conditions. It is by no means certain that this still applies when the surfaces are covered with mica, though some kind of surface barrier may be expected. Whatever its height and shape may be, it cannot be a primary factor governing the surface recombination mechanism. Recombination under steady state conditions clearly involves the acquisition by the surface of electrons and holes at equal rates, accompanied by a redistribution of energy. This must be true for all barriers and in this context it does not matter which are the minority and which the majority carriers.

One of the simplest interpretations of the observed effect could be based on the postulate that surface recombination depends on the presence of recombination centres and that the concentration of such centres, in turn, depends on the capacitively applied field. Recombination centres may arise from adsorbed gas and the applied field could displace the prevailing equilibrium between the surface and the surrounding gas atmosphere. Such changes in the concentration of adsorbed gas have also been inferred from measurements of surface conductance (Low 1955). On the other hand, no more detailed picture of this mechanism can be formulated at present, pending much more elaborate experiments in vacuum and in controlled gas atmospheres. Moreover, the quick response of the system to the applied voltage may be in conflict with the present hypothesis. The following alternative interpretation is therefore favoured.

In the interpretation of bulk phenomena, electrons and holes enter as qualitatively symmetrical concepts and this is a relation which is characteristic of the lattice and its band structure. However, a discussion of surface recombination must probably concern itself with the electronic interaction between the semiconductor and adsorbed foreign layers. It is thus necessary to consider the probabilities of electron loss and electron capture by the ions, atoms or molecules which could constitute these layers. A fundamental asymmetry between holes and electrons then arises, inasmuch as electron capture enlarges the electron cloud associated with the capturing centre, whereas electron loss (hole capture) diminishes it. Electron capture would thus imply a movement of the capturing centre away from the germanium surface. By analogy with the Franck–Condon principle, these transitions should occur in a time too short to allow for any appreciable movement of any atomic nucleus involved. The incident electron is in the neighbourhood of a capturing centre for only a very short time, and electron capture is thus a relatively improbable process. Electron loss leaves the centre in an excited vibrational state and the excess kinetic energy can be communicated to the lattice gradually without time restriction. On this basis, electron capture should be the rate determining process for surface recombination in n-type and p-type material.† We may thus assume that the recombination centres at the surface are almost in dynamic equilibrium with the hole concentration near the surface, but not with the electron concentration. The electron concentration depends on the magnitude of the external field, as applied during the present experiments. If the external plate is positive, the electron concentration at the surface is increased and hence also the rate of electron capture by recombination centres. This would, in turn, lead to an increase of s. Conversely, a negative plate voltage should be associated with a diminution of s, as observed.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Professor R. W. Ditchburn for placing research facilities at our disposal, to Drs. P. C. Banbury and J. W. Mitchell for valuable discussions, and to the Admiralty for their support and for permission to publish this paper.

---

† The authors are indebted to Dr. J. W. Mitchell, University of Bristol, for pointing this out and drawing their attention to this interpretation.

REFERENCES

BRATTAIN, W. H., 1951, *Semi-conducting Materials*. Ed. H. K. Henisch (London: Butterworths Scientific Publications), p. 37.

BRATTAIN, W. H., and BARDEEN, J., 1953, *Bell Syst. Tech. J.*, **32**, 1.

HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81**, 835.

HENISCH, H. K., REYNOLDS, W. N., and TIPPLE, P. M., 1955, *Physica*, **20**, 1033.

KEYS, R., and MAPLE, T. G., 1954, *M.I.T. Lincoln Lab. Quarterly Progress Report*, 1st February 1954.

LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 10.

MANY, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.

REYNOLDS, W. N., 1953, *Proc. Phys. Soc.* B, **66**, 899.

SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York: Van Nostrand), p. 321.

The Dielectric Constant of Aqueous Ionic Solutions

By V. I. LITTLE
Bedford College, London

MS. received 20*th December* 1954, *and in amended form* 21*st January* 1955

Abstract. The variation with concentration of the dielectric constant of aqueous ionic solutions at low frequencies is explained in terms of the radial ordering of solvent molecules about the ions, and by the effects of temporarily linked ion pairs.

---

§ 1. Introduction

In order to explain the high-frequency dispersion of the ionic conductivity of dilute solutions of potassium chloride, it was proposed (Little and Smith 1955) that each ion in solution was surrounded by a number of concentric shells of solvent molecules, each molecule oriented radially by the field due to the ion. It was shown that the application of an external field to such an arrangement would give rise to a reaction field at the centre of the system reducing the net field acting on the ion. At sufficiently high frequencies, owing to the finite relaxation time of the molecules constituting the shell, the reaction field would decrease and the conductivity would increase.

It is the purpose of this paper to derive an expression for the variation of the dielectric constant with concentration of aqueous ionic solutions, using the hypotheses employed in the work mentioned above, namely (i) the existence of radially oriented spherical shells of solvent molecules surrounding each ion, and (ii) the dielectric constant everywhere within the ordered region of a shell system is unity.

§ 2. The Electric Polarizability of a Radially Polarized Spherical Shell

In figure 1 a solvent molecule at P is shown in its equilibrium position under the influence of the radial field $\mathscr{F}$ due to a positive monovalent ion at O, and the field F (parallel to OA) arising from the application of an external field E. If the dielectric constant over the whole of the spherical shell of radius r is taken as unity, it follows that

$$\mathscr{F} = e/r^2 \qquad \ldots\ldots(1)$$

where e is the charge on the ion, and

$$F = E + 4\pi I/3 \qquad \ldots\ldots(2)$$

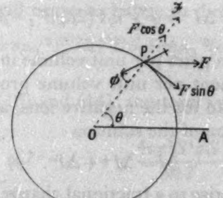


Figure 1.

where I is the mean electric moment per unit volume throughout the whole medium produced by an externally applied field E. If μ_0 is the true dipole moment of the solvent molecules within the medium, the effective moment μ of the molecule at P is given approximately by the Langevin formula

$$\left.\begin{aligned}\mu &= \mu_0[\coth(\mu_0\mathscr{F}/\boldsymbol{k}T) - \boldsymbol{k}T/\mu_0\mathscr{F}]\\ &= \mu_0\mathscr{L}(\mu_0\mathscr{F}/\boldsymbol{k}T)\end{aligned}\right\} \qquad \ldots\ldots(3)$$

where $\boldsymbol{k}$ is Boltzmann's constant, and T is the absolute temperature.

The angle of inclination ϕ of the molecule at P to the radial direction is given by

$$\tan\phi = F\sin\theta/(\mathscr{F} + F\cos\theta)$$

and provided $\mathscr{F} \gg F$ it follows that

$$\cos\phi = 1 \quad \text{and} \quad \sin\phi = F\sin\theta/\mathscr{F}. \qquad \ldots\ldots(4)$$

The contribution of the molecule at P to the moment in the direction OA is given by

$$dm = \mu(\cos\theta\cos\phi + \sin\theta\sin\phi).$$

With equation (4) this becomes

$$dm = \mu(\cos\theta + F\sin^2\theta/\mathscr{F}). \qquad \ldots\ldots(5)$$

If n molecules are uniformly distributed over the surface of the shell of radius r, it is possible to integrate equation (5), thus

$$m = \int_0^{\pi} \tfrac{1}{2}\, n\mu\,\{\sin\theta\cos\theta + F\sin^3\theta/\mathscr{F}\}\,d\theta = 2n\mu F/3\mathscr{F}$$

whence the electric polarizability of the shell may be written

$$p = 2n\mu/3\mathscr{F}. \qquad \ldots\ldots(6)$$

§ 3. The Variation of the Dielectric Constant for Very Dilute Solutions

According to Debye (1929) the contribution of a molecule of moment μ_0 to the moment per unit volume of a liquid subjected to an extrenal field E is given by $\mu_0^2F/3\boldsymbol{k}T$ where F is defined according to equation (2). Should n molecules be locked into a spherical distribution about an ion there will be a change in the moment of the liquid given by

$$dM = (2n\mu/3\mathscr{F} - n\mu_0^2/3\boldsymbol{k}T)F. \qquad \ldots\ldots(7)$$

If each ion succeeds in orienting i concentric shells of n_i molecules, the change in moment per unit volume will be given by

$$q\sum_i dM_i = \Delta I^+ = q\sum_i n_i F(2\mu_i/\mathscr{F}_i - \mu_0^2/\boldsymbol{k}T)/3 \qquad \ldots\ldots(8)$$

where q is the number of ion pairs per unit volume in the solution, and ΔI^+ is the change in the electric moment per unit volume produced by the positive ions. A similar expression will hold for the negative ions, and we may write for the net change in electric polarization of the solution

$$\Delta I = \Delta I^+ + \Delta I^-.$$

The change ΔI will give rise to a fractional change $\Delta\epsilon$ in the dielectric constant given by

$$\Delta\epsilon = 4\pi\Delta I/E. \qquad \ldots\ldots(9)$$

Writing $\Delta I^+ = qJ^+F$ and $\Delta I^- = qJ^-F$ (to express equation (8) for positive and negative ions respectively), equation (9) becomes

$$\Delta\epsilon = 4\pi q[J^+ + J^-]F/E$$

and since

$$\epsilon = 1 + 4\pi I/E$$

$$\Delta\epsilon = 4\pi q(\epsilon+2)(J^+ + J^-)/3 \qquad \ldots\ldots(10)$$

using equation (2).

In practice J^+ and J^- are both negative, whence, writing $\epsilon = \epsilon_0 - \Delta\epsilon$ where ϵ_0 is the dielectric constant of pure water, substitution into equation (10) gives

$$\Delta\epsilon = \frac{4\pi q(\epsilon_0+2)(J^+ + J^-)}{3 + 4\pi q(J^+ + J^-)}. \qquad \ldots\ldots(11)$$

The error involved in assuming ϵ in equation (10) to be ϵ_0 is never large enough however to necessitate the use of equation (11).

For sufficiently dilute solutions equation (10) gives the rate at which the dielectric constant of an ionic solution is reduced as the number of ion pairs per unit volume is increased.

§ 4. THE FORMATION OF ION PAIRS AT HIGHER CONCENTRATIONS

For any concentration there will be a finite probability of finding a negative ion lying at a distance between S and $S+dS$ from a given position ion. This probability for one ion pair per unit volume is given by the Boltzmann expression $4\pi S^2 dS \exp(-U/\boldsymbol{k}T)$ where U is the potential energy of the negative ion.

If there are q ion pairs per unit volume, and if it is assumed that the concentration is such that the potential energy of the negative ion is almost exclusively due to the field of the positive ion at a distance S, then the number of pairs of ions separated by distances between S and $S+dS$ is given by

$$n = q^2 4\pi S^2 dS \exp[e^2/\epsilon S\boldsymbol{k}T] \qquad \ldots\ldots(12)$$

where ϵ is the mean dielectric constant of the medium.

We now make the assumption that when a pair of ions of opposite sign are separated by a distance lying between S and $S+dS$, they exhibit dipolar characteristics in the sense that a temporary linkage takes place, and the pair exists as a single unit with a dipole moment μ_0' given by

$$\mu_0' = eS. \qquad \ldots\ldots(13)$$

The contribution of n linked pairs of this kind, per unit volume, to the electric polarization will be given by $\Delta I' = n\mu_0' \mathscr{L}(\mu_0' F/\boldsymbol{k}T)$, and provided $\mu_0' F \ll \boldsymbol{k}T$, $\Delta I' = n\mu_0'^2 F/3\boldsymbol{k}T$. This will cause a change in dielectric constant given by

$$\Delta\epsilon' = 4\pi n\mu_0'^2 F/3\boldsymbol{k}TE = 4\pi n\mu_0'^2(\epsilon+2)/9\boldsymbol{k}T \qquad \ldots\ldots(14)$$

as in the previous section.

Combining equations (10) and (14), the total variation in dielectric constant $\delta\epsilon$ becomes

$$\delta\epsilon = \Delta\epsilon + \Delta\epsilon' = \frac{4\pi}{3}(\epsilon+2)\left[q(J^+ + J^-) + q^2\frac{4\pi e^2 S^4 dS}{3\boldsymbol{k}T}\exp\left(\frac{e^2}{\epsilon S\boldsymbol{k}T}\right)\right] \qquad \ldots\ldots(15)$$

in which equations (12) and (13) have been used to substitute for μ_0' and n in equation (14).

§ 5. THE EVALUATION J^+ AND J^-

The x-ray analysis of water by Morgan and Warren (1938) and the interpretation of this work by Pople (1951) led the author in a previous paper (Little and Smith 1955) to assume that in ionic solutions the radii of the shells of water molecules surrounding an ion may be considered to be essentially the same as the radii of the shells of nearest neighbours surrounding a given water molecule in pure water. The error in assuming these radii to be 2, 4, 6, 8, 10... Å, will not lead to any widely diverging results for the purposes of the present paper.

The radius of the outermost shell surrounding an ion will be determined by the relative sizes of the forces acting on a molecule in the shell, due to the central ion, and to an adjacent molecule. For water these forces are almost equal in a shell of radius 10 Å, assuming that the dielectric constant is unity. Beyond this distance the normal tetrahedral structure of the water will become re-established, and the dielectric constant will rapidly regain its normal high value.

Pople's calculation for pure water yielded 4, 11, and 22 for the number of molecules occupying shells one, two and three. Little and Smith (1955) have shown that some evidence exists for the numbers 4, 12 and 27 to represent the population of the first three shells surrounding an ion. Unfortunately, there is no direct evidence concerning the number of molecules in the fourth and fifth shells which can contribute to the net polarization of the system. The dispersion effect of the conductivity at very high frequencies could be explained satisfactorily in terms of the contribution of the 27 effective molecules in the third shell, assuming that the relaxation times associated with the molecules in the first and second shells were very small. This assumption however implied that any molecule in the fourth or fifth shell which could contribute did so, and its contribution was reckoned as if it were occupying a position in the third shell. It is concluded therefore that the number of effective—as distinct from actual—molecules in shells four and five is very small.

In table 1 are given the values of the contributions to J for the first five shells, computed from equation (7), remembering that $dM = J_i F$ and using equations (1) and (3).

Table 1

r (Å)	Shell	$J_i \times 10^{24}$	n_i
2	1	$-25{\cdot}8$	4
4	2	$-22{\cdot}9$	12
6	3	$-18{\cdot}8$	27
8	4	$-14{\cdot}8$	n_4
10	5	$-11{\cdot}7$	n_5
∞	∞	$-9{\cdot}0$	0

The third column was calculated using the values

$$\mu_0 = 1{\cdot}8 \times 10^{-18} \text{ for water, } \boldsymbol{k} = 1{\cdot}38 \times 10^{-16},\ T = 300^\circ\text{K},\ e = 4{\cdot}8 \times 10^{-10}\,\text{e.s.u.}$$

It will be noted that the value for J_∞ is only a little smaller than one-half of the value for J_3, which is another sound reason for the rapid diminution of n_i as i approaches the larger values.

The figures available for n_i refer only to potassium chloride solution, and it is assumed in this case that both potassium and chlorine ions possess the same shell structures. Some evidence in favour of this assumption is that these ions have the same mobilities in aqueous solutions. Assuming that n_5 is zero, the value of J^+ and J^- is obtained by adding the products of the figures in columns 3 and 4 of table 1, and doubling ; thus

$$J^+ + J^- = -(1776 + 2q \, . \, 3n_4) \times 10^{-24}. \qquad \ldots\ldots(16)$$

§ 6. THE DIPOLE MOMENT OF LINKED PAIRS

It has been assumed that a radial ordering of solvent molecules takes place about a positive and negative ion in solution. If this ordering extends to a radius r about the ion, it follows that when two ions of opposite sign but with similar shell structures approach each other, there will be an ordering of the molecules between them as soon as the field at the centre of the pair is equal to the field at a distance r from either. From this it may be concluded that two ions will be linked by a bridge of directed solvent molecules when their separation is $\sqrt{8}r$. The dipole moment of such a pair would be $\sqrt{8}re$ minus the net moment of the extra molecules constituting the bridge. If the ions approach to within a distance $2r$, then the axial field intensity at either end of the dipole will be less than it is for a single ion at a similar distance. The effect of this will also be to cause some reduction in the dipole moment of the doublet. For simplicity it is assumed that linkages occur when the outermost shells of two ions of opposite sign come into contact, that is, when the separation between charges is 20 Å. A tolerance of 0·5 Å chosen experimentally determines the limits between which dipole formation of the type described can take place.

With these assumptions it is possible to express equation (15) numerically. Instead of q, the number of ion pairs in unit volume, it is more convenient to use the substitution

$$q = 6{\cdot}03C \times 10^{20} \qquad \ldots\ldots(17)$$

where C is the concentration of salt in gramme molecular weights per litre. Using this substitution, equation (16), and putting S and dS equal to 20 and 0·5 Å respectively, equation (15) becomes

$$\delta\epsilon = 34 \times 10^3 C^2 - [370 + 6n_4]C, \qquad \ldots\ldots(18)$$

a parabolic form in which the coefficient of C^2 is determined by the formation of dipole type linkages, and the coefficient of C by the shell structure about the ions in solution. It will be noted that the parabolic form is not affected by the method employed in determining the coefficient of C^2.

At concentrations of the order 0·01 normal, the distance separating one linked pair from another is about 200 Å. This is about ten times the separation of the charges constituting the pair, and only five times the distance between the extremities of the shells of the pair. It may be expected therefore that strong dipole interaction should take place for concentrations of 0·01 normal and above, which would reduce the contribution provided by the rotation of the linked pairs.

In figure 2 equation (18) has been plotted for values of C below 0·014 normal, and for values of n_4 between 0 and 12. In addition the experimental values for potassium chloride solutions over the same range of concentrations obtained by Pechhold (1927) are plotted for comparison purposes. The agreement is not good, but this may be due to the large experimental errors in the results for potassium chloride, and to the assumption that the chlorine and potassium ions have similar shell structures.

A curve could be made to fit the experimental results if it is assumed that the chlorine ion shell structure is more extensive than that of the potassium ion, but lack of adequate experimental evidence for this makes such a calculation pointless. It will be clear from equation (18) that any modifications to the coefficient of C^2 will tend to alter the width of the parabola at the point where it cuts the concentration axis drawn through the point $\epsilon = 80$, without changing its initial slope. Hence, in order to obtain a better fit with the experimental results, the value of $J^+ + J^-$ in equation (16) must be increased, which implies a more extensive shell structure than has been assumed.

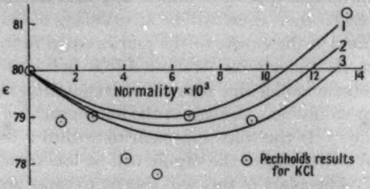


Figure 2. The variation of the dielectric constants of KCl solutions with concentration. Curves 1, 2, 3 are plots of equation (18) for $n_4 = 0$, 6 and 12 respectively.

The upward trend of the curve at higher concentrations shown in figure 2 will be reduced by the effect of interaction between linked pairs, but it is not possible at present to give more than the rough estimate already given of the concentration at which this effect would become apparent.

§ 7. THE EXPERIMENTAL RESULTS OF PECHHOLD, AND MILICKA AND SLAMA

The measurement of the dielectric constant of electrolytic solutions presents a very difficult problem owing to the effect of relatively large values of the conductivity. The first successful attempt to measure the dielectric constant of sodium chloride solutions was made by Fürth (1924), using an original method in which the couple acting on an ellipsoid suspended in the solution was measured as a function of the applied electric field and the dielectric constant. Fürth reported that the dielectric constant of sodium chloride decreased to a minimum value as the concentration increased, and eventually increased to values in excess of the value at zero concentration. Later, Pechhold (1927), using Fürth's technique, carried out an investigation on the chlorides of hydrogen, lithium, sodium and potassium. These results are reproduced in figure 3. Unfortunately, the accuracy of these results for potassium chloride is not good, owing to the greater difficulty with this salt arising from its large conductivity at the concentration employed. However, there is little doubt that the variations of dielectric constant are of the same general form as those for the more accurately reproducible curves.

In 1931 a further investigation was carried out by Milicka and Slama on the halides of sodium and numerous other salts. Unfortunately, potassium chloride was not dealt with, but the sodium chloride curve was essentially the same as the results obtained by Pechhold.

It is interesting to note that although all the curves of figure 3 exhibit similar characteristics, the curve for potassium chloride is much shallower than the curves for the other salts of the group. The explanation of this effect may be found if it is assumed that the shell structure around the smaller ions of lithium and sodium is more strongly bound and more extensive than the structure around potassium. Some information concerning the nature of these structures can be obtained from the initial slopes of the curves in figure 3. These are determined by the coefficient of C in equation (18). Substitution of equation (17) into equation (15) gives the expression

$$\text{Initial slope} = 0{\cdot}21(J^{+}+J^{-})\times 10^{24};$$

hence the values of J^{+} and J^{-} may be tabulated for the curves in figure 2 and are given in table 2.

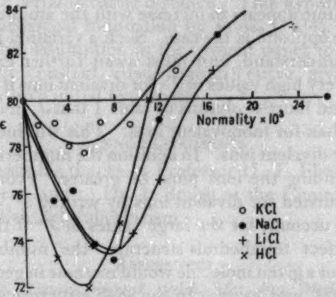


Figure 3. Pechhold's results (1927) for the variation of the dielectric constant with concentration of solutions KCl, NaCl, LiCl and HCl.

Table 2

	KCl	NaCl	LiCl	HCl
$(J^{+}+J^{-})\times 10^{24}$	−2250	−5400	−5800	−11800

Assuming that J^{+} and J^{-} are equal for KCl, the value of J^{+} for the other positive ions may be obtained by simple subtraction, and yield the results in table 3.

Table 3

	Cl^{-}	K^{+}	Na^{+}	Li^{+}	H^{+}
$J\times 10^{24}$	1125	1125	4300	4700	11000

By definition J is the electric polarizability per ion per unit volume and is given by

$$J = \sum_i J_i n_i = \sum_i \frac{\mu_0}{3}\left[2\mathscr{L}\left(\frac{\mu_0\mathscr{F}}{kT}\right)\Big/\mathscr{F}_i - \frac{\mu_0}{kT}\right] n_i. \qquad \ldots\ldots(19)$$

If the variation of the dielectric constants of other than monovalent salts is investigated, the variation being plotted against concentration in normality, i.e. gramme equivalent weights per litre, the initial slopes of these curves would be given by the expression

$$\text{Initial slope} = 0{\cdot}21\left[\frac{1}{v}J^{+}+J^{-}\right]\times 10^{-24}$$

where the positive ion valency is v, and the negative ion is univalent. Milicka and Slama have investigated the chlorides of copper, barium, aluminium, lead and ferrous and ferric iron. Applying the above formula to these results and using the value for Cl^- in table 2, the polarizabilities per ion per unit volume of these elements have been found and are tabulated in table 4.

Table 4

	Fe^{2+}	Cu^{2+}	Ba^{2+}	Pb^{2+}	Al^{3+}	Fe^{3+}
$J \times 10^{24}$	20000	18000	37000	160000	48000	72000

The results in tables 3 and 4 are subject to large errors, due to the fact that the slopes of the experimental curves of Pechhold, and of Milicka and Slama, could not be accurately assessed at zero concentration. The only conclusion which can be drawn at the moment is that for divalent positive ions, the polarizability per ion per unit volume appears to increase with the atomic number, whereas for monovalent ions the opposite is the case. Such a variation in the case of divalent ions is difficult to understand, and must await further experimental evidence. To account for the very high values of J^+ for divalent ions it is clear from equation (19) and the third and fourth columns of table 1 that J_i must fall off more slowly with i for divalent than for monovalent ions. This will indeed be the case since $\mathscr{F}_i$ will be greater for divalent ions. In addition the numbers of molecules effective in the shells surrounding the ions must be greater. From the data in table 5 which has been calculated for divalent ions by writing $\mathscr{F}_i = 2e/r_i^2$ it will be seen that it is possible to account for the large values of J^+ if the numbers n_i are increased. In a perfect tetrahedral structure, the numbers of molecules in successive shells about a given molecule would increase in geometrical progression, namely 4, 12, 108, 324, and so on. In the shell structure about divalent ions, it is not improbable that with radial packing these numbers could be exceeded. The values of n_i, the number of molecules in a shell effectively contributing to the effect, could therefore assume fairly large values. For instance, the values of J^+ for copper, barium and lead in table 4 could arise from suitable combinations of values of n_i, together with the corresponding values of J_i taken from table 5.

Table 5

r (Å)	Shell	$J_i \times 10^{24}$	n_i max	J_i max $\times 10^{24}$
2	1	26·4	4	105
4	2	25·0	12	300
6	3	22·8	36	820
8	4	20·1	108	2170
10	5	17·3	324	5600
12	6	14·0	972	13600
14	7	12·9	2916	37600
16	8	11·6	8748	101500
∞	∞	9	0	0

In order to account for the very large values of J for barium and for lead, it may be assumed that the innermost shell is absent, thus reducing any screening effect which this shell could produce, and increasing the numbers of effective molecules in the outer shells.

§ 8. CONCLUSION

In its present form the theory throws some light on the manner in which water molecules are oriented by ions in solution. A detailed investigation of the variation of the dielectric constant of aqueous ionic solutions at very low concentrations would provide valuable information concerning the extent of the field due to the ions, and possibly the effective size of the larger ions.

The fact that Debyes' theory of dielectrics as applied above can be used to obtain some measure of quantitative agreement with experiment is encouraging, in spite of the known limitations of this theory. More exact numerical data could be obtained by using Onsager's theory (1936), but it was desired in the present case to attempt an explanation of the dielectric constant variation with concentration in terms of the simplest, and most generally known model. The assumption in § 4 concerning the formation of temporary dipoles is a simplified version of Bjerrum's hypothesis (1926 a, b) that positive and negative ions could pair together temporarily and exhibit dipole characteristics, if their separation fell below a certain critical distance. This hypothesis was later given a more rigorous mathematical foundation by Fuoss (1934). It is interesting to note that the use of this assumption in the present paper to explain the positive slope of the (dielectric constant, concentration) curve, gives some measure of corroboration of Bjerrum's initial hypothesis.

ACKNOWLEDGMENTS

The author expresses his thanks to Professor H. T. Flint and Dr. N. Hill of the Physics Department, and to Dr. J. Smith of the Chemistry Department of Bedford College for their helpful criticism.

REFERENCES

BJERRUM, N., 1926 a, *K. Danske Vidensk. Selsk. Mat. Fys. Medd.*, **7**, No. 9; 1926 b, *Ergebn. Exakt. Naturw.*, **6**, 125.
DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalog Co.).
FUOSS, R., 1934, *Trans. Faraday Soc.*, **30**, 967.
FÜRTH, R., 1924, *Phys. Z.*, **25**, 676.
LITTLE, V. I., and SMITH, V., 1955, *Proc. Phys. Soc.* B, **68**, 65.
MILICKA, VON O., and SLAMA, A., 1931, *Ann. Phys., Lpz.*, **8**, 663.
MORGAN, J., and WARREN, B. E., 1938, *J. Chem. Phys.*, **6**, 666.
ONSAGER, L., 1936, *J. Amer. Chem. Soc.*, **58**, 1486.
PECHHOLD, R., 1927, *Ann. Phys., Lpz.*, **83**, 427.
POPLE, J. A., 1951, *Proc. Roy. Soc.* A, **205**, 163.

The Effect of Field-Dependent Mobilities on the Diffusion Theory of Rectification

By P. T. LANDSBERG
University of Aberdeen, Scotland

Communicated by T. E. Allibone; MS. received 26th March 1954, *and in final form 4th December* 1954

Abstract. It is shown that, if the mobility and diffusion coefficient of the current carriers are subjected to a much smaller fractional change as a function of position in the barrier layer than is the electric field, then many results of the general theory of barrier layer rectifiers remain essentially unchanged.

IN recent experimental work on copper oxide rectifier characteristics, Lees (1953) confirmed a predicted connection between the current–voltage and capacitance–bias voltage characteristics. The analysis was based on a theory of these characteristics which is 'general' in that it does not presuppose a definite theoretical law for the space charge distribution in the barrier layer. Nevertheless some rather crude assumptions are still involved even in this work: (*a*) Einstein mobility relation $eD/v=\boldsymbol{k}T$, (*b*) field-independent mobility v, (*c*) single pronounced potential energy hump V_{m} in the barrier such that $V_{\mathrm{m}}v/eD \gg 1$, (*d*) negligible effects due to minority carriers, (*e*) validity of the microscopic image force law (i.e. barrier must be at least about 20 Å thick), (*f*) neglect of space charge due to current carriers, (*g*) plane geometry. We wish to show that Lees' conclusions are unaffected if (*b*) is omitted and (*a*) replaced by the much weaker assumption (*a*′) that for given applied voltage, temperature, etc., the fractional variation of mobility v and diffusion coefficient D in the barrier $(0<x\leqslant d)$ is much less than the fractional variation of the electric field. If (*a*) and (*b*) are retained, and one replaces assumption (*c*) by stipulating a certain presumed distribution of fixed space charge, and assumption (*e*) by stipulating neglect of the image force, one can obtain the older theories of Mott and Schottky.

Dropping assumptions (*a*) and (*b*), integration of the diffusion equation of rectification is still possible, and yields

$$N(x)\exp\left(-\int_x^d \frac{v}{eD}\frac{dV}{dy}\,dy\right)-\int_x^d \frac{j}{eD}\exp\left(-\int_y^d \frac{v}{eD}\frac{dV}{dz}\,dz\right)dy=N(d), \qquad (1)$$

as may be verified by differentiating. j is here the current density from the metal to the semiconductor, e and N algebraic charge and volume density of the current carriers respectively. v, D and v/eD are allowed to depend on x. Assuming that close to the metal electrode a plane $x=a$ exists such that

$$V(a)=0, \qquad N(a)=C, \qquad \ldots\ldots(2)$$

independent of applied voltage, it follows that $C=N(d)=N(a)$ (Landsberg 1951 a, equation (7)). Also (1) yields

$$j=\frac{eN(d)}{J}\left\{\exp\left(-\int_a^d \frac{v}{eD}\frac{dV}{dx}\,dx\right)-1\right\}, \qquad \ldots\ldots(3)$$

where

$$J=\int_a^d \frac{1}{D}\exp\left(-\int_x^d \frac{v}{eD}\frac{dV}{dz}\,dz\right)dx. \qquad \ldots\ldots(4)$$

The coordinate of the plane at which the exponent in (4) reaches its maximum is the coordinate at which V reaches its maximum, and has in previous papers (Landsberg 1951 b) been denoted by $x=b$. By (a'),

$$\int_b^d \frac{v}{eD}\frac{dV}{dx}\,dx \sim \frac{v(b)}{eD(b)}\,[V(d)-V(b)]. \qquad \ldots\ldots(5)$$

The integral J is, by (c), of a form to which Laplace's approximation

$$\int_a^d f(x)\exp[g(x)] \sim \left(-\frac{2\pi}{g_{\mathrm{m}}''}\right)^{1/2} f_{\mathrm{m}}\exp g_{\mathrm{m}} \qquad \ldots\ldots(6)$$

can be applied, where the subscript m denotes values at the maximum of $g(x)$. We shall also use (Landsberg 1951 b)

$$V_{\mathrm{D}}-V(b)=\frac{e^2}{2\epsilon b}, \qquad -V''(b)=\frac{e^2}{2\epsilon b^3}, \qquad \ldots\ldots(7)$$

where ϵ is the dielectric constant of the semiconductor, and V_{D} is the potential energy difference between the top of the barrier, in the absence of image force and tunnel effect, and the bottom of the conduction band in the interior of the semiconductor when no current is flowing. This is also the zero of the potential energy. There are obvious modifications in this definition when conduction is mainly by holes. Using equations (3) to (7),

$$j=j_2y^{3/2}e^y\left\{\exp\left(-\int_a^d\frac{v}{eD}\frac{dV}{dx}\,dx\right)-1\right\}\exp\left[\frac{v(b)}{eD(b)}V(d)\right], \qquad \ldots\ldots(8)$$

$$j_2\equiv(2/\pi)^{1/2}\epsilon N(d)\,\frac{D^2(b)}{v(b)}\exp\left[-\frac{v(b)}{eD(b)}V_{\mathrm{D}}\right], \qquad \ldots\ldots(9)$$

$$y\equiv ev(b)/2\epsilon bD(b). \qquad \ldots\ldots(10)$$

The values of v and D to be used in (9) and (10) are those which would be obtained in the limit of zero field, so that j_2 is independent of voltage on the present model. The product of the last two factors in (8) tends to unity for reverse voltages, and one obtains the same form for the *reverse* current–voltage relation as in the previous paper. If one bears in mind that the capacitance–bias voltage relation does not involve the mobility or the diffusion coefficient explicitly, it follows that assumption (b) can be dispensed with, when working out the theory of Lee's experiments. Most of the $\boldsymbol{k}T$ terms that occur must then, however, be reinterpreted as $eD(b)/v(b)$.

If the Einstein mobility relation is assumed, equations (8), (9) and (10) do not involve the field dependence of the mobility. Thus, for barriers to which the assumptions stated at the beginning apply, the field dependence of the mobility cannot affect the current–voltage curves to any marked degree.

Burgess (1953) pointed out that if, *throughout the barrier*,

$$v(x)\frac{dV}{dx}=A, \qquad eD(x)/v(x)=\boldsymbol{k}T, \qquad \ldots\ldots(11)$$

where A is independent of x, then the diffusion equation can be integrated *exactly*. This is due to the fact that the introduction of (11) into (4) yields

$$j=\frac{e}{A}\exp\left[-\frac{V(d)}{\boldsymbol{k}T}\right]\left\{\exp\left[-\frac{V(a)}{\boldsymbol{k}T}\right]-\exp\left[-\frac{V(d)}{\boldsymbol{k}T}\right]\right\}, \qquad \ldots\ldots(12)$$

and there is then no need to appeal to approximate methods of integration.

If, however, (11) is introduced into the approximate result (8), the formula obtained by Burgess is not reproduced. The reason is that the present approximate integration does not apply to the case represented by (11). For, by assuming that (11) is valid throughout the barrier, one is in fact implying that

$$\lim_{x\to b} D(x) = \lim_{x\to b} v(x) = \begin{cases} +\infty & (0<x\leqslant b \\ -\infty & (b\leqslant x<d). \end{cases} \qquad \ldots\ldots(13)$$

Thus, since v and D are rapidly varying functions near $x=b$, assumption (a') is violated, and (8) does not apply.

If one wishes to take into account the fact that $v\,dV/dx=A$ *for strong fields*, it is clear that Burgess sets up the problem approximately, by assuming this condition to be valid throughout the barrier, but performs the integration exactly. The treatment leading to (8), on the other hand, enables one to set up the problem exactly, but the integration is then carried out approximately. The advantage of the second method is that it gives a more accurate account of the conditions near the potential maximum.

REFERENCES

BURGESS, R. E., 1953, *Proc. Phys. Soc.* B, **66,** 430.
LANDSBERG, P. T., 1951 a, *Proc. Roy. Soc.* A, **206,** 463; 1951 b, *Ibid.*, **206,** 477.
LEES, J., 1953, *Proc. Phys. Soc.* B, **66,** 622.

The Growth of an Electron Avalanche Retarded by its Own Space Charge

By G. FRANCIS
Clarendon Laboratory, Oxford

Communicated by A. von Engel; MS. received 30th November 1954, *and in revised form* 24*th January* 1955

Abstract. In this paper calculations are given showing how the ionization coefficient α, and the total multiplication in the avalanche, vary with the distance travelled when the positive ions formed have an appreciable effect. The time taken for an avalanche to develop is also worked out. Plane parallel geometry is considered, and two important cases are treated, namely constant field at the electrodes, or constant potential across them. A simple and exact solution is obtained for constant field, and an approximate solution for constant potential. The latter shows that α may decrease or increase depending on the initial conditions, and that it tends finally to a constant value. The equations are general and thought to be of wide application.

§ 1. Introduction

ELECTRONS driven by an electric field through a gas collide with gas molecules and, by virtue of their random motion, ionize some of them. The number of electrons increases by the factor $e^{\alpha x}$ as the swarm moves a distance x in field direction ; more generally by $\exp(\int \alpha dx)$ if the field is not uniform. α is Townsend's ionization coefficient. This formula is usually applied only when the total number of ions and electrons is too small to distort the applied electric field. But in many physical systems large numbers are produced giving fields comparable with and often equal to that applied. If we say that n_0 initial electrons are multiplied to $n_0 \exp(\int \alpha dx)$ it is clear that large numbers can be produced in two ways. A few initial electrons may be enormously multiplied by moving a large distance in the field (e.g. lightning, sparks in long gaps), or a relatively large initial number may be multiplied by some reasonable factor, say 10 or 100, by moving a short distance. This latter can occur when a secondary process gives rise to a series of successive avalanches, the last few of which have a large n_0, e.g. current pulses in low-frequency electrodeless discharges.

The initial electrons, colliding with gas molecules, produce more electrons which move rapidly in field direction leaving behind them a tail of almost stationary positive ions. These slow down the electrons and retard multiplication until finally so many ions have been created that virtually no further ionization is possible.

The electrons at the head of the avalanche thus move in a field which varies both in space and time. The purpose of this paper is to calculate, on the simplest assumptions, how this field varies in space, and by using the appropriate value of α to calculate $\int \alpha dx$, and hence the multiplication up to any point x. The physical assumptions, discussed below, are admittedly open to criticism on a few points, and it is also questionable whether in strongly non-uniform fields it is permissible to use the concept of α at all, because the electrons are not in equilibrium with the field. However, they have the merit of leading to a manageable mathematical treatment which gives general results.

It should be emphasized that these calculations refer only to multiplication in a single avalanche, no secondary processes being invoked. Here we consider only a single process (ionization by electron collision) in the gas and the results are therefore equally applicable whether the gas is contained between metal electrodes or insulating walls. Radial diffusion is also neglected.

§ 2. The Physical Assumptions and Their Validity

The physical assumptions are illustrated in figure 1. We assume that (i) Initially n_0 electrons start at a plane and move in a uniform field E_0. (ii) As the avalanche develops the electrons travel in front in an extremely thin layer whose thickness can be neglected. This layer contains no ions, which are all left behind. (iii) The motion of the electrons is determined by the field X_x at the centre of the electron layer when it is in the position x. Thus X_x determines both the value of α at this point, α_x, and the drift velocity of the cloud of electrons. (iv) To a first approximation α_x is proportional to X_x. (v) It is permissible to use the normal Townsend equation for multiplication even in non-uniform fields. (vi) No secondary processes affect appreciably the growth of the one avalanche being considered.

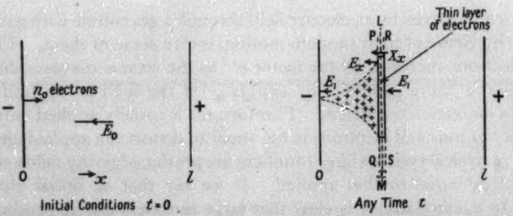


Figure 1. The physical picture. E_x=field at trailing edge of electron layer, X_x=field at centre of electron layer. Unit cross section considered (the width of the electron layer is exaggerated for clarity). The curve illustrates only the growing density of positive ions, and does *not* imply lateral diffusion, see § 4.

Assumptions (ii) and (iii) are valid over a very wide range. Two factors tend to widen the electron layer : diffusion, and the screening effect of the electrons themselves, in that those at the leading edge of the layer suffer little retardation from the positive ions and so move further ahead. This screening effect becomes serious only when the number of electrons is extremely large, so that a large difference exists between the fields at the two edges of the layer (E_x and E_1, figure 1). In other words the layer begins to widen only when the space charges almost cancel the applied field, and this marks nearly the end of the avalanche. Similarly, until this stage has been reached the field at all points in the layer is not very different (e.g. ± 20%) from the field at its centre, which can thus be taken to determine the behaviour of the whole cloud.

The spreading by diffusion has to be compared with the time taken for the avalanche to progress, i.e. with the drift velocity of the electrons. The drift velocity depends upon X/p, diffusion upon $1/p$, and so for avalanches occurring at comparable values of X/p the spreading of the layer will become less at higher pressures. Take a typical numerical example: an avalanche in hydrogen with $X/p=30$ takes 2×10^{-7} sec to cross a 2 cm gap, in which time the electrons have diffused rather less than 1 mm from their centre, if the gas pressure is 10 mm Hg. Thus the layer is never more than 2 mm wide. In general the approximation of a thin layer is reasonable at pressures between 1 and 10 mm Hg, and extremely good at higher pressures. In long gaps the spreading by diffusion becomes less in proportion to the gap length.

It is easily seen that the layer consists of electrons alone, and that there is negligible overlapping with the ions. Consider the cloud of n electrons passing the point x. All n electrons will occupy the layer between x and $x+dx$, but the only ions in that layer will be $dn=n\alpha dx$, a negligible number, compared with the electrons.

The assumption that α is proportional to the local field is tolerably good provided X/p is not large, for example, less than about 50 for hydrogen, nitrogen or air. This is one of the few assumptions that becomes more valid towards the end of the growth of the avalanche, when local fields become low.

Assumption (vi) is arbitrary but reasonable, and the validity of using α (assumption (v)) is questionable only in very non-uniform fields.

§ 3. THE BASIC EQUATIONS

The conditions are as shown in figure 1 at time t. α refers to the general value, when the electron cloud has reached a point x.

We consider throughout unit area of cross section. The number dn of electrons and ions formed when a cloud of n electrons moves from x to $x+dx$ is

$$dn=n\alpha dx \qquad \ldots\ldots(1)$$

giving the number of electrons n_x at x as

$$n_x=n_0\exp\left(\int_0^x \alpha dx\right) \qquad \ldots\ldots(2)$$

where n_0 is the initial number at $x=0$.

Since $\alpha=f(x)$ this formula includes all variations of α due to the changing local field.

Applying Gauss' theorem between the planes LM, RS (figure 1) we find

$$X_x=E_1-2\pi n_0\mathbf{e}\exp\left(\int_0^x \alpha dx\right). \qquad \ldots\ldots(3)$$

Applying the theorem between the planes PQ, RS (figure 1) we find

$$E_x=E_1-4\pi n_0\mathbf{e}\exp\left(\int_0^x \alpha dx\right). \qquad \ldots\ldots(4)$$

If α is proportional to X_x, then

$$\alpha=bX_x \qquad \ldots\ldots(5)$$

where b is a constant.

We may now consider two sets of conditions of physical importance. In one the field at the boundary of the gap is constant, i.e. the charge on the electrode or wall is constant during the growth of the avalanche. This would be true in certain stages of the lightning discharge. In the other the potential across the gap is kept constant, which would be so in most laboratory experiments. As the avalanche develops the field distribution alters, and hence the field at the boundary.

We have then two alternative equations :

For a given boundary field

$$E_1 = E_0 = \text{constant}. \qquad \ldots\ldots(6)$$

For a given applied potential V,

$$V = \int_0^l E_x\,dx$$

which, with equation (4), gives

$$V = E_1 l - 4\pi n_0 \mathbf{e} \int_0^x \left\{ \exp\left(\int_0^x \alpha dx \right) \right\} dx. \qquad \ldots\ldots(7)$$

§ 4. SOLUTION WHEN THE APPLIED FIELD IS CONSTANT

Equations (1) to (6) now apply.

In equation (3) substitute for X_x from (5), and differentiate with respect to x:

$$\frac{1}{b}\frac{d\alpha}{dx} = -\alpha 2\pi n_0 \mathbf{e} \exp\left(\int_0^x \alpha dx \right). \qquad \ldots\ldots(8)$$

Differentiating this equation and dividing the result by (8),

$$\frac{d^2\alpha/dx^2}{d\alpha/dx} = \alpha + \frac{1}{\alpha}\frac{d\alpha}{dx}. \qquad \ldots\ldots(9)$$

Substitution of $p = d\alpha/dx$ leads to a first-order linear equation

$$\frac{d}{d\alpha}\left(\frac{p}{\alpha}\right) = 1. \qquad \ldots\ldots(10)$$

Integration, with the boundary conditions

$$x = 0,\ \alpha = \alpha_0,\ \int \alpha dx = 0 \text{ and } \left(\frac{d\alpha}{dx}\right)_{x=0} \text{ as given by (8),} \qquad \ldots\ldots(11)$$

leads to

$$\alpha = \frac{B}{1 + P\exp(Bx)} \qquad \ldots\ldots(12)$$

where

$$B = \alpha_0 + 2\pi n_0 \mathbf{e} b$$

$$P = \frac{2\pi n_0 \mathbf{e} b}{\alpha_0} = \tfrac{1}{2}\frac{4\pi n_0 \mathbf{e}}{E_0} \qquad \ldots\ldots(13)$$

i.e. one-half the ratio of initial space charge field to applied field. Both B and P are positive constants.

This equation shows that as $x \to \infty$, $\alpha \to 0$ which is physically sensible.

The multiplication M is obtained simply as follows:

$$\int_0^x \alpha dx = Bx + \log \frac{1+P}{1+P\exp(Bx)}. \qquad \ldots\ldots(14)$$

The constant of integration being defined by (11),

$$M=\exp\left(\int_0^x \alpha dx\right)=\frac{1+P}{P+\exp(-Bx)}. \qquad \ldots\ldots(15)$$

Here again a simple check shows that as $x\to\infty$

$$M\to \text{constant}=1+1/P \qquad \ldots\ldots(16)$$

which is also physically sensible.

These results also show, as we would expect, that the rate at which α decreases and the avalanche comes to a standstill depends on n_0, the initial number of electrons and also upon the relation between α and the local field (factor b). A large number n_0 of initial electrons and a strong field (large α_0, hence large B) both produce a rapid fall in α. Figure 2 shows a numerical example, calculated from equation (12).

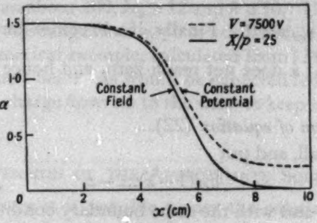


Figure 2. A numerical example, for H_2, illustrating how the assumption of constant potential (equation 24) leads to a less rapid fall in α than a constant field at the electrodes (equation 12). $p=30$ mm Hg, 10 cm gap, and either $V=7500$ volts (constant potential) or $X/p=25$, constant at electrodes. $n_0=10^5$. Note how, when constant potential is applied, α does not become zero but tends to $\alpha=2/l$.

§ 5. SOLUTION WHEN THE APPLIED POTENTIAL IS CONSTANT

When the applied potential is constant equations (1) to (5), and (7) apply.

In equation (3) substitute for X_x from (5), multiply by l and subtract equation (7):

$$\frac{l}{b}\alpha-V=2\pi n_0 e\left[2\int_0^x\left\{\exp\left(\int_0^x \alpha dx\right)\right\}dx-l\exp\left(\int_0^x \alpha dx\right)\right]. \qquad \ldots\ldots(17)$$

Differentiating with respect to x

$$\frac{l}{b}\frac{d\alpha}{dx}=2\pi n_0 e\left\{\exp\left(\int_0^x \alpha dx\right)\right\}(2-l\alpha). \qquad \ldots\ldots(18)$$

This equation has an important significance which will be discussed in § 11. Differentiating again with respect to x, and dividing the resulting equation by (18) and putting $p=d\alpha/dx$, we obtain

$$\frac{d}{d\alpha}\left(\frac{p}{l\alpha-2}\right)=\frac{\alpha}{l\alpha-2}. \qquad \ldots\ldots(19)$$

Here we substitute $l\alpha=e^u$ (implying, as is physically consistent, that α is always positive), and integrate with respect to u using the same boundary conditions as before, except that $(d\alpha/dx)_{x=0}$ is now given by (18).

Introducing the new variable

$$z=\frac{\exp(u_0)-\exp(u)}{\exp(u_0)-2}=\frac{l(\alpha_0-\alpha)}{l\alpha_0-2} \qquad \ldots\ldots(20)$$

and putting

$$\exp(u_0)-2=k, \text{ and } C'=Pl\alpha_0 \qquad \ldots\ldots(21)$$

leads to:

$$\frac{dz}{(1-z)[kz+C'-2\log(1-z)]}\doteq\frac{dx}{l} \qquad \ldots\ldots(22)$$

and the solution of the problem lies in the integration of the left-hand side. This cannot be done rigorously but good approximations can be found over appreciable ranges, depending on the initial conditions.

Range of values of z.

The space charge reduces α as the avalanche develops; thus from (20) z is of constant sign and if, as in most physical applications, $l\alpha_0>2$, then z is always positive.

Initially when $\alpha=\alpha_0$, $z=0$. Finally, if α becomes zero, $z=\frac{l\alpha_0}{l\alpha_0-2}=1+\frac{2}{k}$.

Later we shall see that α does not reach zero, and hence z does not extend over this whole range.

Approximate integration of equation (22).

If we assume z small, and put

$$\log(1-z)\simeq -z \qquad \ldots\ldots(23)$$

integration is possible and with the same boundary conditions leads to

$$\alpha=\alpha_0-\frac{(l\alpha_0-2)}{l}\left\{\frac{\exp\{l\alpha_0(1+P)x/l\}-1}{1/P+\exp\{l\alpha_0(1+P)x/l\}}\right\}. \qquad \ldots\ldots(24)$$

A few simple checks show that when $x=0$, $\alpha=\alpha_0$, $d\alpha/dx=2\pi n_0 eb(2-l\alpha_0)/l$, agreeing with (18), and also that α decreases as x increases, which is consistent with the physical picture.

It can also be seen that when $x/l\to 1$ (if the solution is valid up to that value, see § 6), $\alpha\to 2/l$: also if $l\alpha_0<2$, then α increases as the avalanche develops. This is discussed in § 11.

§ 6. RANGE OF VALIDITY OF THE APPROXIMATE SOLUTION

The approximation (23) is itself accurate to within 10% for values of z up to 0·2. But the total error may be much less than this because $\log(1-z)$ is only one of three terms in the denominator of (22), all of them positive.

k depends on $l\alpha_0$ and C' on n_0, i.e. both depend only on the initial conditions. If $l\alpha_0$ is large k is large and this has two favourable results: not only is the error in the above approximation very much reduced, so that the solution can be used up to larger values of z, but the total range of z (0 to $1+2/k$) is much reduced. Thus when k is very large the solution (24) may be used over the whole range of z, i.e. from $\alpha=\alpha_0$ to $\alpha=0$. C' can have a range of values from almost zero up to that determined by the largest value of n_0.

The largest value of C' likely to have any practical application occurs, let us say, when the field due to the n_0 electrons is about 10% of the applied field, leaving a multiplication of about 10 necessary before the avalanche ceases. This leads to $C'\leqslant l\alpha_0/10$. C' has a much smaller effect on the expression than k, although again a large C' reduces the error resulting from the approximation.

Some practical values.

Consider, for example, hydrogen in a plane parallel gap to which starting potential V_s is applied. Now the whole treatment is valid only for those values of X/p for which we can regard α as proportional to the field. For hydrogen this is $X/p \leqslant 50$, roughly ; from the curve of V_s plotted against pl (pressure × gap length) this occurs when $pl \geqslant 12$, and then $l\alpha_0 \geqslant 3$, i.e. $k \geqslant 1$.

The most unfavourable conditions for using this approximate solution in hydrogen are $pl \simeq 12$, $l\alpha_0 \simeq 3$ when the maximum error is 9% for range $z = 0$ to $0{\cdot}25$, $C' \simeq 0$, i.e. n_0 very small, and 6% for range $z = 0$ to $0{\cdot}25$, $C' = l\alpha_0/10$, i.e. $n_0 \simeq 10^{10}$.

Better conditions occur at larger values of pl. For example, when $pl \geqslant 45$ then $l\alpha_0 \geqslant 6$, i.e. $k \geqslant 4$. The solution may be used then with less than 6% error up to $z = 0{\cdot}3$, or 9% error up to $z = 0{\cdot}4$ when $C' \simeq 0$, with the errors becoming 1% less when n_0 becomes large.

When pl is greater than about 200, $l\alpha_0 \geqslant 12$ and $k \geqslant 10$. Under these conditions the solution applies over the range from $z = 0$ to $z \simeq 1$ with less than 5% error.

Figure 2 shows a numerical example, calculated from (24), which is compared with that given by constant field. As would be expected from physical reasoning α falls less rapidly when charge flows on to the walls to keep the potential constant.

§ 7. EXTENSION OF THE APPROXIMATE SOLUTION

In circumstances where it is possible to use the solution (24) already given only over a restricted range say from $z = 0$ to $z = f$, it would be useful to have another solution over an adjacent range.

This can be achieved by returning to the left-hand side of equation (22). If we put

$$z = f + y \text{ and } w = y/(1-f) \qquad \text{......(25)}$$

we get an equation, with w as the variable, of the same form as (22) which can be integrated by using the same kind of approximation, i.e. for w small put $\log(1-w) \simeq -w$.

The boundary condition is chosen so that both the formulae, i.e. equation (24) and the result we are trying to find, will be satisfied by the same values of α and x at the point $z = f$. Thus when $z = f$, $w = 0$ and x/l is given by putting $z = f$ in (24).

This leads finally to

$$\alpha = \alpha_0 - \frac{l\alpha_0 - 2}{l}\left\{f - (1-f)\frac{q - G\exp\{(p+q)x/l\}}{p + G\exp\{(p+q)x/l\}}\right\} \qquad \text{......(26)}$$

where in the previous notation:

$$q = kf + C' - 2\log(1-f) = l\alpha_0(f+P) - 2f - 2\log(1-f)$$

$$p = k(1-f) + 2 = l\alpha_0 - f(l\alpha_0 - 2)$$

$$p + q = k + 2 + C' - 2\log(1-f) = l\alpha_0(1+P) - 2\log(1-f)$$

and

$$G = \left[\frac{q(1-f)}{1 + l\alpha_0 f/C'}\right]^{\frac{p+q}{C'+l\alpha_0}}$$

$$= \left\{\frac{(1-f)[l\alpha_0(1-f) - 2f - 2\log(1-f)}{1 + f/p}\right\}^{\frac{1-2\log(1-f)}{l\alpha_0(1+P)}} \qquad \text{......(27)}$$

§ 8. RANGE OF VALIDITY OF THE EXTENDED SOLUTION

A treatment as outlined in § 6 shows that the permitted range of w is greater than that of z for the same percentage error. A numerical example, similar to that given for z (e.g. $f=0{\cdot}3$, C' negligible, $k \geqslant 4$) shows that the error is less than 6% for $w \leqslant 0{\cdot}4$. Remembering that $y=w(1-f)$ we find that this extension roughly doubles the original range of z provided this original range was less than about 0·3, i.e. if (24) was valid for the range $z=0$ to f then (26) will be valid for the range $z=f$ to $2f$, and the range of w will be from 0 to $f/(1-f)$ provided $f \leqslant 0{\cdot}3$. The total error when both formulae are used will be the sum of the errors involved in each.

Further extensions of this kind can obviously be made, but each one requires more tedious labour which hardly seems justified.

§ 9. MULTIPLICATION GIVEN BY THE APPROXIMATE SOLUTIONS FOR α

Equation (24) and its extension (26), giving α as $f(x)$, can both be put in the form

$$\alpha = R - S\left[\frac{\exp(mx/l) - T}{\exp(mx/l) + V}\right] \qquad \ldots\ldots(28)$$

where R, S, T, V, m are constants. The method of integration is the same for both equations, but with different boundary conditions. For the first range (equation (24) z small) $\int \alpha\, dx = 0$ when $x=0$, and integration gives the multiplication M as

$$M = \exp\left(\int_0^x \alpha dx\right) = \left[\frac{1}{1 + P\exp\{l\alpha_0(1+P)x/l\}}\right]^{(l\alpha_0-2)P/l\alpha_0} \exp[\{l\alpha_0(1+P) - 2P\}x/l]. \qquad \ldots\ldots(29)$$

If this equation is valid only for the range $z=0$ to $z=f$, and the multiplication is M_0 when $z=f$, further multiplication is given by using equation (26) with the appropriate boundary conditions: from (24) the value of x/l corresponding to $z=f$ is

$$\frac{x_0}{l} = \frac{1}{C' + l\alpha_0}\log\frac{1 + l\alpha_0 f/C'}{1-f} = \frac{1}{l\alpha_0(1+P)}\log\frac{1+f/P}{1-f}. \qquad \ldots\ldots(30)$$

Substitution of this value in (29) gives M_0.

Integration of (26) then gives for the next range the total multiplication M as

$$M = \exp\left(\int_0^x \alpha dx\right) = M_0\left[\frac{\exp\{(p+q)x_0/l\} + p/G}{\exp\{(p+q)x/l\} + p/G}\right]^{(l\alpha_0-2)(1-f)/p}$$
$$\times \exp\left\{\left[l\alpha_0 + (l\alpha_0 - 2)\left\{(1-f)\frac{q}{p} - f\right\}\right]\frac{x-x_0}{l}\right\}. \qquad \ldots\ldots(31)$$

§ 10. THE TIME TAKEN FOR AN AVALANCHE TO DEVELOP

The solutions for α, equations (12), (24) and (26) can be written generally

$$\alpha = f(x) \qquad \ldots\ldots(32)$$

which combined with (5) gives

$$bX_x = f(x). \qquad \ldots\ldots(33)$$

But X_x the field at the centre of the electron cloud determines the drift velocity dx/dt of the electrons. If some simple relation between them is known equation

(33) can be integrated giving the time taken for the avalanche to progress to a distance x.

For the molecular gases (H_2, N_2, O_2 and air) we have approximately

$$V_{\text{drift}} = dx/dt = L + NX_x \qquad \ldots\ldots(34)$$

where L, N are constants.

Using equations (33) and (12) in (34),

$$\frac{dx}{dt} = L + \frac{N}{b}\,\frac{B}{1+P\exp(Bx)} \qquad \ldots\ldots(35)$$

which can be integrated with the boundary condition $t=0$ when $x=0$, giving finally

$$t = \frac{b}{Lb+NB}\left\{x + \frac{N}{Lb}\log\frac{NB+Lb\{1+P\exp(Bx)\}}{NB+Lb(1+P)}\right\}. \qquad \ldots\ldots(36)$$

The use of equations (24) and (26) for the variation of α at constant potential leads to exactly similar relations between time and distance, only with different constants.

§ 11. THE FINAL EQUILIBRIUM VALUE OF α WHEN CONSTANT POTENTIAL IS APPLIED

One rather surprising result emerges from equations (24) and (26) giving α as $f(x)$; this is that α never becomes zero, but as $x/l \to 1$ α tends to a limit

$$\alpha = 2/l \qquad \ldots\ldots(37)$$

(see, for example, figure 2).

This is true whatever the initial conditions; the above equations show that if $l\alpha_0 > 2$, α decreases from α_0 to $2/l$, while if $l\alpha_0 < 2$ then α increases from α_0 to $2/l$. Moreover, this final result does not depend upon any approximations made in the solution of the equations, because equation (18), derived without any approximations from the basic equations, gives

$$\frac{l}{b}\,\frac{d\alpha}{dx} = 2\pi n_0 e\left\{\exp\left(\int_0^x \alpha\,dx\right)\right\}(2 - l\alpha) \qquad \ldots\ldots(18)$$

and hence $\quad d\alpha/dx = 0$ when $\alpha = 2/l$. $\qquad \ldots\ldots(38)$

Since all the other terms are positive the sign of $d\alpha/dx$ depends directly upon $(2 - l\alpha)$. Thus α will first approach and, if the multiplication is sufficient, finally reach and remain constant at the value $2/l$.

These arguments imply that, with the physical assumptions made, the avalanche never stops but multiplication continues indefinitely, with a constant $\alpha = 2/l$, after a certain stage of the development has been reached. Such an unexpected result needs further consideration.

Suppose that figure 3 represents the field distribution at some instant during the development of the avalanche. The discontinuity $2h$ represents the change in field across the electron layer; its height represents the number of electrons in the layer and hence the multiplication, while α is given by the mean field in the layer, i.e. the mid-point of the discontinuity.

Now assume that α remains constant: we shall try to find its value. Let the electrons move forward a distance ξ, when the dotted line will show the new field distribution. Conditions to be satisfied are: (i) The multipilcation, and hence the height of the discontinuity, will have increased by the factor $e^{\alpha\xi}$ but because α is constant the mid-point remains at the same level. With the dimensions as shown on the diagram this gives

$$h+d=he^{\alpha\xi}\ ,\ \text{i.e. } d=h(e^{\alpha\xi}-1). \qquad \ldots\ldots(39)$$

(ii) The field at both electrodes must be the same and so the whole diagram is lifted up by the height d. (iii) The curve between x and ξ, given by the ion distribution produced as the electrons move forward with constant α, must be of the form $2\,he^{\alpha y}$, from $y=0$ to $y=\xi$. (iv) The total area under the curve, representing the potential, must remain constant. The shaded sections added, and the dotted section subtracted from the original area must balance.

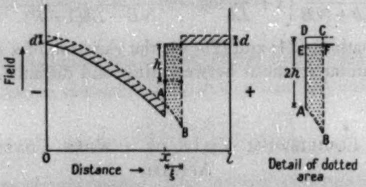


Figure 3. The field distribution produced by an advancing avalanche when constant potential is applied. This diagram shows how, after a certain stage, α can remain constant. The curved part on the left represents the field distribution due to stationary ions formed in earlier stages of the avalanche.

Now the areas added are easily seen to be: xd on the left, $d[l-(x+\xi)]$ on the right.

$$\text{The total area added} = d(l-\xi). \qquad \ldots\ldots(40)$$

The area subtracted (see small side diagram) is best regarded as

$$\text{ABCD} - \text{CDEF} = 2h\int_0^{\xi} e^{\alpha y}\,dy - \xi\,d.$$

$$\text{Total area subtracted} = \frac{2}{\alpha}h(e^{\alpha\xi}-1)-\xi\,d \text{ which, using (39),}$$

$$= d\left(\frac{2}{\alpha}-\xi\right). \qquad \ldots\ldots(41)$$

Equating (40) and (41) gives

$$\alpha = 2/l. \qquad \ldots\ldots(42)$$

Hence all the necessary conditions are satisfied and clearly α can remain constant i.e. $2/l$. Multiplication thereafter continues and increases as $e^{\alpha\xi}$.

Physically this may be explained as follows. The development of an avalanche in a gap having a constant potential across it produces two effects: the ions reduce the field in the electron cloud in the manner already explained, but at the same time the motion of charge alters the field distribution in the gap, and more charge flows on to the electrodes to keep the potential constant. This increases the field on the electrons. When the initial field is low ($l\alpha_0<2$) this latter effect predominates and α increases; if the initial field is large ($l\alpha_0>2$) the reverse happens. When the avalanche has developed in space until at some point $l\alpha=2$

the two effects are equal. The electrons then move in constant field and multiplication continues at a constant rate. Of course there comes a time when the number of electrons becomes so enormous that they no longer move in anything approximating to a plane, and this treatment is no longer valid. Even so we have deduced enough to disprove the generally accepted idea that the avalanche comes to a standstill, and also to show that much larger multiplications and concentrations of electrons can be produced than might previously have been imagined.

§ 12. SOME PRACTICAL APPLICATIONS

This analysis would be expected to apply chiefly to the development of a single avalanche in a long gap between parallel electrodes, and at pressures above 1 mm Hg where the electrons would remain in a relatively thin layer, and for values of X/p less than 50 for molecular gases, or less than 30 for rare gases. If starting potential or less is applied this corresponds roughly to pl greater than 15 or 30 respectively. The application of the formulae becomes more convenient and accurate at larger values of pl. Since secondary processes are neglected the treatment is only applicable in circumstances where the secondary mechanism coefficients are small, or where there is a relatively long time delay between primary and secondary processes, for example when further avalanches are started by ions hitting the cathode. A sufficient number of ions (usually 100 to 1000) must be produced by the first avalanche which will have proceeded a great distance by the time the ions reach the cathode: the second avalanche will have then little effect upon the development of the first. A photoelectric effect at the cathode, although acting rapidly, often gives a low γ because many photons are lost by geometry. The treatment would apply to the initial stage of spark breakdown between parallel plates, where cloud chamber photographs show the growth of an avalanche before a streamer is formed. Here the potential is kept constant.

The initial stage of the lightning discharge from a negative cloud is believed to consist of a series of choked avalanches, with $X/p \simeq 40$. Considering the size of the cloud relative to the extent of an avalanche, and remembering that the charge on the cloud does not change during the growth of any one avalanche, the conditions can be regarded as a uniform constant field, and the treatment of § 4 should apply.

§ 13. CONCLUSIONS

By making simple assumptions it has been found possible to calculate how Townsend's ionization coefficient α varies with distance x as an avalanche develops due to the retarding effect of the positive ions left behind, and to calculate the multiplication of the avalanche as a function of the distance and determine the time taken for it to develop.

If the field at the boundary of the gap is constant the equations derived are simple and exact (granted the physical assumptions). The assumption of constant potential results in more difficult equations which can be solved approximately in steps of increasing complexity. In many circumstances one such step only is necessary and the equation can be solved to an accuracy within 10%; two steps would be sufficient for most problems, and would give an accuracy to about 20% or better.

One important conclusion results from the assumption of constant potential: α may decrease or increase with distance depending on the initial values of α and n_0, and does not tend finally to zero as might be expected but to a constant finite value. This implies that multiplication continues indefinitely, at least until the physical assumptions of this treatment no longer apply.

ACKNOWLEDGMENTS

This work has arisen out of research being done at the Clarendon Laboratory supported by British Electricity Authority. I am grateful to Professor Lord Cherwell for the facilities of the Laboratory and to the Research Council of B.E.A. for a grant and for permission to publish this paper. I want expecially to thank Dr. A. von Engel for much encouragement, and Dr. Handel Davies, Christ Church, for valuable help with § 5 of the paper.

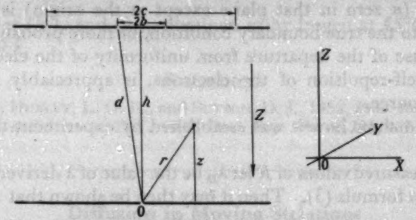
RESEARCH NOTES

A Note on the Diffusion in a Gas of Electrons from a Small Source

By L. G. H. HUXLEY and R. W. CROMPTON
Department of Physics, University of Adelaide

MS. received 23*rd December* 1954

In recent studies of the motions of electrons in gases based on the methods introduced by Townsend (Crompton, Huxley and Sutton 1953 and references therein), diffusion chambers of the design indicated in the figure were used.



Electrons enter the diffusion chamber through a small hole O in the cathode and move towards the anode in a steady state of motion under the action of a uniform electric field Z which gives them a mean drift speed W. The stream spreads by diffusion with a coefficient K. The differential equation describing the distribution of the electron concentration n is $\nabla^2 n = 2\lambda \partial n/\partial z$ where $2\lambda = W/K$.

The source O was made small, in contrast with the earlier designs of apparatus, in order to obtain solutions for n, when O is in the origin of coordinates and the cathode the XOY plane, in the form of a series whose $(m+1)$th term is

$$A_m \exp(\lambda z)(\lambda r)^{-1/2} K_{m+1/2}(\lambda r) P_m(\cos\theta). \qquad \ldots\ldots.(1)$$

It was hoped that in practice the series would be found to reduce in effect to a single term, but it does not seem possible to predict *a priori* with any confidence, which term ; consequently recourse must be had to experiment.

The experimental technique for determining $\lambda = W/2K$ consists in measuring the ratio R of the current i_1 to a central insulated portion (radius b) of the anode to the total anode current i_2. The formula which has been used throughout this work to derive λ from R is

$$R = \frac{i_1}{i_2} = 1 - \frac{h}{d}\exp[-\lambda(d-h)] \qquad \ldots\ldots(2)$$

and experimental evidence, as shown below, established that this is beyond doubt the correct formula over a wide range of values of λ.

This formula was first obtained by adopting as a solution that term from equation (1) for which $m=1$, that is to say, a dipole solution which gives

$$n = A_1 \lambda^{-1/2} \exp(\lambda z) \frac{z}{r} \frac{d}{dr}\left[\frac{\exp(-\lambda r)}{r}\right].$$

In order to satisfy the boundary condition $n=0$ over the anode $z=h$ an image source of suitable strength was supposed to exist at the point (0, 0, $2h$), and the electron flux $(-Kdn/dz)_{z=h}$ was then integrated to give the currents i_1 and i_2.

It may be shown, however, that this procedure with a dipole source does not in fact lead to formula (2) but to the following formula :

$$R=1-\left[\frac{h}{d}\right]^3\frac{(1+\lambda d)}{(1+\lambda h)}\exp\left[-\lambda(d-h)\right]. \qquad \ldots\ldots(3)$$

In order to obtain formula (2), which agrees with experiment, it is necessary to suppose that the correct term to adopt from (1) is that corresponding to the pole ($m=0$) for which

$$n=A_0(\lambda)^{-1/2}\exp(\lambda z)\exp(-\lambda r)/r. \qquad \ldots\ldots(4)$$

It would thus appear that either the cathode boundary solution given by the dipole solution (n zero in that plane except at the origin) is not an adequate approximation to the true boundary condition, or more probable, the differential equation, because of the departure from uniformity of the electric field and the possibility of self-repulsion of the electrons, is appreciably inaccurate in the vicinity of the source.

It remains to show how it was established by experiment that formula (2) is correct.

For some measured values of R let λ_D be the value of λ derived from formula (2) and λ_D that from formula (3). Then it may then be shown that

$$h\lambda_D+\frac{1}{d\lambda_D}=h\lambda_P-1+(\text{small terms}) \qquad \ldots\ldots(5)$$

However, by definition, the true λ is $W/2K$ which from general considerations of similitude satisfies the following functional relationship:

$$\lambda=W/2K=p\mathrm{F}(Z/p) \qquad \ldots\ldots(6)$$

where p is the gas pressure at constant temperature.

It follows therefore, from equation (5) that whichever of λ_P or λ_D is the true λ therefore satisfying some relationship (6), then the other cannot accurately satisfy such a relationship.

In practice the quantity λ has been written $\lambda=20{\cdot}15\,p\,(Z/p)/k_1$ where k_1 is given the interpretation

$$k_1\propto\frac{\text{electron temperature}}{\text{temperature of the gas}}.$$

It therefore follows from equation (6) that k_1 is a function of Z/p only ; consequently, if $(k_1)_P$ and $(k_1)_D$ correspond to λ_P and λ_D, then if one is a function of Z/p only, the other is not.

The following table of experimental results shows clearly that $(k_1)_P$ does not change as p is changed, if Z/p is held unchanged. On the other hand $(k_1)_D$ shows a pressure dependence with variations much greater than the experimental error which was about 2%. A more extensive example of the fact that $(k_1)_P$ is a function Z/p only, is given in the paper by Crompton and Sutton (1952). It is found, moreover, that the same quantitative functional dependence of $(k_1)_P$ upon Z/p is obtained in pieces of apparatus with different dimensions and with the anode divided by cross cuts instead of by circular cuts.

Z/p	p	R	$(k_1)_P$	$(k_1)_D$
0·5	1·5	0·556	5·14	6·85
	2	0·653	5·07	5·96
	3	0·775	5·15	5·77
1·0	1·5	0·592	9·05	11·5
	2	0·690	9·01	10·4
	3	0·813	9·10	10·0
2·0	1	0·546	14·3	19·3
	1·5	0·671	14·4	16·7
	2	0·761	14·4	16·1
5·0	0·75	0·559	25·6	33·7
	1	0·637	26·1	31·3
	1·5	0·770	26·3	29·4

Table of values of $(k_1)_P$ and $(k_1)_D$ obtained in hydrogen at 15°c in an apparatus for which $b/h=0·5$.

References

Crompton, R. W., Huxley, L. G. H., and Sutton, D. J., 1953, *Proc. Roy. Soc.* A, **218**, 507.
Crompton, R. W., and Sutton, D. J., 1952, *Proc. Roy. Soc.* A, **215**, 467.

Diffusion in Moving Striations

By V. D. FARRIS[†]
Physics Department, Queen's University, Belfast

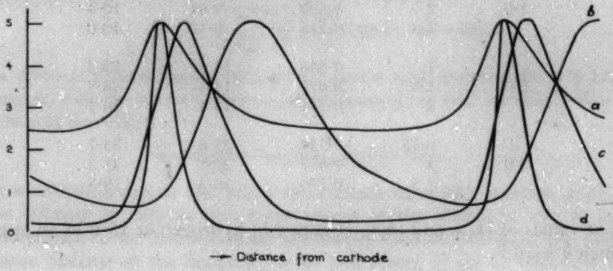
Communicated by K. G. Emeléus; MS. received 14th March 1955

The positive column of a discharge at pressures of the order of 1 mm and above often consists of striations moving with high speed towards the cathode. These have been investigated electrically, using Langmuir probes, by Pupp (1932, 1935). The data published by him are not extensive, and besides being subject to the errors which arise in use of probes in steady discharges, may be liable to further errors from the presence of oscillations. It is known, however, that they account for some of the features of the spectra (Sloane and Minnis 1935). The present Note is a short account of an attempt (Farris 1954) to develop a diffusion theory of the striations, and to test it from Pupp's measurements.

Pupp (1935) has summarized his results for a discharge in argon at a pressure of 1 mm in a tube 4·8 cm in diameter, as charts of instantaneous contours of electron concentration n, electron temperature T, and space potential. The variations of T and n along the axis of the tube are shown by curves a and b in the figure. The corresponding space-charge distribution can be found by applying Poisson's equation to the space–potential contours, assuming that the discharge is symmetrical about the axis. It proves to be considerable only near where the electron temperature has its maximum, the maximum space-charge density being equivalent to an excess N of $1·5\times10^7$ singly charged positive ions per cm³.

Now at the Research Laboratories, General Electric Company Ltd., Wembley.

Pupp does not give absolute values for n, but from the values of n found in other discharges its maximum value was probably about 10^{10} cm^{-3}. Comparison with N of Pupp's curves for n, normalized roughly in this way, indicates that there was little departure from plasma conditions in the column.



Axial discharge data for moving striations. a, temperature T, unit 4800°K; b, concentration n, units arbitrary; c, rate of ionization from diffusion equation s/n, unit $0{\cdot}35 \times 10^4$ ions cm^{-3} sec^{-1}; d, rate of ionization from ionization equation s/n, unit 10^5 ions cm^{-3} sec^{-1}. Distance between peaks, 17·5 cm. Striations moving from right to left.

If it is now assumed that the small difference between the ion and electron concentrations has the same effects axially as radially, and that mean values of mobilities and diffusion coefficients may be used, the theory of a stationary positive column may be readily generalized to include changes of concentration axially and with time. Eliminating the current and field from the usual conservation and current density equations gives

$$s = \frac{\partial n}{\partial t} - D\nabla^2 n \qquad \ldots\ldots(1)$$

where s is the rate of production of ions per unit volume per second, and D the Schottky ambipolar diffusion coefficient, given to a close approximation by (Loeb 1939)

$$D = \boldsymbol{k}Tb/\boldsymbol{e} \qquad \ldots\ldots(2)$$

where b is the positive ion mobility.

Examination of Pupp's graphs for n and T shows that the former varies with radius r approximately as $J_0(2{\cdot}4r/R)$, where R is the tube radius, whilst the radial change in T over a given cross section is much smaller, and will be neglected in what follows.

If $p/2\pi$ is the frequency of the moving striations and $2\pi/k$ their spacing along the direction of the axis z, the curves for n may be analysed into a Fourier series

$$n = \left\{\alpha_0 + \sum_{l=1}^{\infty} \alpha_l \sin\left[l(pt - kz) + \delta_l\right]\right\} J_0(2{\cdot}4r/R). \qquad \ldots\ldots(3)$$

s can be expressed by a similar series,

$$s = \left\{A_0 + \sum_{l=1}^{\infty} A_l \sin\left[l(pt - kz) + \epsilon_l\right]\right\} J_0(2{\cdot}4r/R), \qquad \ldots\ldots(4)$$

and from equation (1) the amplitudes and phases will be connected by the relations

$$A_0 = \alpha_0 D(2{\cdot}4/R)^2 \qquad \ldots\ldots(5a)$$

$$A_l = lp\alpha_l / \sin(\epsilon_l - \delta_l) \qquad \ldots\ldots(5b)$$

$$(\tan\epsilon_l - \tan\delta_l)\{(2{\cdot}4/R)^2 + (lk)^2\}D = lp(1 + \tan\epsilon_l \tan\delta_l). \qquad \ldots\ldots(5c)$$

Actually, since absolute values of n are not known, the analysis will give not s but s/n, as a function of z and r. In carrying out the calculation of s/n from equations (3), (4) and (5) numerically it was found that the Fourier components were negligible for l greater than 5. The values of p and k were 7037 sec^{-1} and 0·363 cm^{-1} respectively. The value of $3{\cdot}7 \times 10^3$ $\text{cm}^2\ \text{sec}^{-1}$ was adopted for D, based on a mean electron temperature of 2×10^4 °K, and values for the mobility given by Massey and Burhop (1952). The resulting curve for s/n along the axis $(r = 0)$ is shown in the figure as curve c. Its peak lies between those of the curves for n and T.

If the ionization were due to collisions between electrons and argon atoms in the ground state, and the electron energy distribution were accurately Maxwellian, the rate of ionization would be given by

$$s/n = 2{\cdot}0 \times 10^7 W^{1/2}(V_1 + 4W/3)\exp(-3V_1/2W) \qquad \ldots\ldots(6)$$

where V_1 is the ionization potential, and W the voltage equivalent of the electron temperature, 450 $\boldsymbol{k}T/\boldsymbol{e}$ (cf. Loeb 1939). The values of s/n calculated in this way are shown in the figure as curve d.

Considering the approximations which have been made, the difference in shape and position of curves c and d is probably not significant. However, the absolute value of the maximum of curve d is 30 times that of curve c. The discrepancy would disappear if the value of s/n calculated from equation (6) were too big either because the values of W given by Pupp were about 40% too high, or because the actual electron energy distribution had a marked deficit of fast electrons compared with the Maxwellian distribution assumed. This might be decided by fresh experiment using modern probe techniques.

ACKNOWLEDGMENTS

The writer wishes to thank the Warren Committee of the Royal Society for financial assistance, and Professor K. G. Emeléus for suggesting this problem.

REFERENCES

FARRIS, V. D., 1954, *Thesis*, Queen's University of Belfast.
LOEB, L. B., 1939, *Electrical Discharge in Gases* (London: Chapman and Hall).
MASSEY, H. S. W., and BURHOP, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: Clarendon Press).
PUPP, W., 1932, *Phys. Z.*, **33**, 844; 1935, *Ibid.*, **36**, 61.
SLOANE, R. H., and MINNIS, C. W., 1935, *Proc. Phys. Soc.*, **47**, 1019.

A Resistance Thermometer for Use at Helium Temperatures

BY D. H. PARKINSON AND L. M. ROBERTS
Radar Research Establishment, Gt. Malvern, Worcs.

MS. received 16*th December* 1954

§ 1. INTRODUCTION

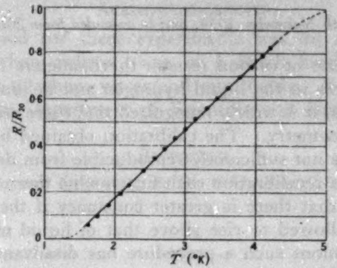
IN a previous note (Parkinson and Quarrington 1954) the resistance–temperature characteristics of leaded brass wires annealed at a number of temperatures between 160°C and 400°C were reported. The composition of the wires was 62% Cu, 36% Zn, 1·73% Pb, 0·08% Ni, and their diameter 0·004 in. The unannealed wire had a broad superconducting type of transition between 4·0°K and 5·4°K and it was found that, except for temperatures between 240°C and 310°C, the resistance characteristic of the wire moved approximately parallel to itself and to lower temperatures as the annealing temperature was increased. It was remarked that two or more wires suitably annealed and joined in series could be used as a thermometer over the whole range 1·5°–4·2°K. Such thermometers have been made and used in this laboratory, but it is difficult to choose annealing temperatures and lengths of wire which produce a characteristic free from marked irregularities. The characteristic would be expected to improve, however, as the number of wires used is increased. An equivalent result has been obtained by annealing a single length of wire in a suitable temperature gradient, the resulting thermometer having a nearly linear characteristic between 1·5°K and 4·2°K. The method and results are given below.

§ 2. EXPERIMENTAL PROCEDURE

The temperature gradient for the annealing process was produced as follows· A slug of iron (2 in. long × $\frac{3}{4}$ in. diameter), which fitted into an electric furnace, was hard soldered to one end of a copper rod (O.F.H.C. 24 in. long × $\frac{1}{2}$ in. diameter) having a narrow groove cut along one side. Hard soldered into this at equal intervals were six thermocouples, the leads to which were protected by thin glass sleeves and fitted snugly into the groove. These leads were brought out at the other end of the rod. Concentric with the rod was a heavily lagged 1 in. Yorkshire copper tube, one end of which also entered the furnace, the other being water cooled. When the furnace was controlled at about 400°C the thermocouples showed that the temperature gradient along the rod was nearly linear, and covered the range 120°–400°C. The leaded brass wire was wound evenly (3 turns/inch) on to that part of the rod whose temperature had been found to lie between 200°C and 385°C except between 225°C and 310°C where the minimum length of wire was used. The furnace temperature was then maintained at the required value for 3½ hours, previous experiments having shown this to be sufficient time to complete most of the changes produced by annealing. The rod was then removed from the furnace and allowed to cool to room temperature. After annealing the wire was wound on to a small former and its resistance measured at helium temperatures, using a constant current of 50 microamperes.

§ 3. RESULTS

In the figure the reduced resistance R/R_{20} (R_{20} is the resistance of the wire at 20°K) has been plotted against temperature in the helium vapour pressure range and a straight line has been drawn through the points. There are two regions where the points depart appreciably from the line and, if a linear characteristic were assumed, the measured temperatures would be in error by $-0{\cdot}02$°K around 2·8°K and $+0{\cdot}02$°K around 1·8°K. For rough measurements such errors can be neglected and the wire constitutes an extremely useful linear thermometer. Where greater accuracy is required the wire can be calibrated in a conventional manner. Its resistance is, of course, dependent on current and magnetic field.



Reduced resistance–temperature characteristic of leaded brass wire.

This work establishes a method by which a suitable sample of leaded brass wire can be converted into a sensitive and nearly linear thermometer covering the helium vapour pressure range of temperature.

ACKNOWLEDGMENTS

The authors wish to thank Mr. T. W. Bayliss, who constructed the annealing apparatus and annealed the wire. This note is published by permission of the Chief Scientist, Ministry of Supply.

REFERENCE

PARKINSON, D. H., and QUARRINGTON, J. E., 1954, *Proc. Phys. Soc.* B, **68**, 644. (It should be noted that in figures 1 and 2 of this paper the ordinate scale should read from 0 to 1·0, not 0 to 10.)

Carbon Resistor Thermometers

By F. E. HOARE, J. C. MATTHEWS and B. YATES
The Physics Laboratories, University of Leeds

MS. received 10*th November* 1954, *and in amended form* 20*th January* 1955

Although the use of carbon resistor thermometers for the measurement of temperatures in the liquid hydrogen and helium ranges has become quite common, it is widely recognized that there are intrinsic difficulties in this form of thermometry. The calibration obtained by plotting resistance against temperature is not sufficiently reproducible from day to day to do away with the necessity of a recalibration each time such a thermometer is used. We have been informed that there is greater constancy if the temperature of the thermometer is not allowed to rise above that of liquid nitrogen at any time, but for many applications such a procedure has disadvantages, even in those cases where it is possible.

During the course of experiments in the liquid hydrogen and helium ranges extending over several months we have found that plotting the reduced resistance R_T/R_F against temperature for one such thermometer, kept isolated from the refrigerant, the reproducibility of the curve is much improved although the thermometer is allowed to warm up to room temperatures between runs. R_T is the resistance at the temperature T and R_F that at some fixed temperature, conveniently taken as the normal boiling point of hydrogen or helium. Our method of resistance measurement is to use a bridge circuit, operated under constant current conditions, and to regulate the current supplied to the bridge. In experiments at liquid and solid hydrogen temperatures the current was regulated by a meter but ultimately in the helium range the potential drop across a standard resistance in the supply circuit to the bridge was kept constant, with a noticeable increase in the reproducibility of the results. Owing to our experimental arrangements it is not easily possible to make measurements over the two temperature ranges successively without allowing the thermometer to warm up in between. Consequently the continuity of the calibration curves over the range from 1·6 to 20°K cannot be tested from our results.

We have found that writing the equation proposed by Clement and Quinnell (1952) in the slightly modified form

$$\log(R_T/R_F)+\frac{K}{C+\log(R_T/R_F)}=A+\frac{B}{T},$$

where C is a further constant, a satisfactory representation of our results may be obtained. With suitable choice of constants the greatest difference between the calculated and observed resistance in the range 10–20°K was 0·1% (176 results over a period of three months); in the range 1·6–4·2°K the greatest divergence was 0·3% (348 results over a period of four months). The corresponding temperature divergences are of the same order. This agreement in calibration over several weeks suggests that in many experiments a single measurement at or near a fixed temperature could subsequently be substituted for the tedious re-calibration over the whole range of temperature which is necessary each time resistance rather than reduced resistance is employed.

This note is written in the hope that other workers who have suitable data will test whether they give good reproducibility in the curves for temperature variation when treated in the manner outlined above.

ACKNOWLEDGMENTS

The resistor used for these experiments was an Allen-Bradley nominal 56 ohm $\frac{1}{2}$ watt carbon resistor, one of a batch kindly supplied to us by Dr. J. R. Clement to whom we extend our thanks. B. Y. wishes to express his gratitude to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCE

CLEMENT, J. R., and QUINNELL, E. H., 1952, *Rev. Sci. Instrum.*, **23**, 213.

LETTERS TO THE EDITOR

The Effect of Adsorbed Air on Photoconductive Layers of Lead Sulphide

Experiments have been carried out in these laboratories to investigate the change in properties of evaporated layers of lead sulphide (about 1 micron thick and with a crystallite size of the order of $\frac{1}{10}\mu$) which have adsorbed varying quantities of air.

A layer of suitably sensitized lead sulphide was placed in a vacuum chamber designed so that Hall constant and photoconductive measurements could be made *in situ*. The layer which was initially prepared under vacuum was exposed to air at atmospheric pressure. When equilibrium was reached the following quantities were measured at room temperature : i(μA) the current through the unilluminated layer which had 10v across its electrodes, Δi(μA) the increase in current when the layer was illuminated by a tungsten lamp, p the number of current carriers per cubic centimetre, μ the Hall mobility of these carriers in cm^2 v^{-1} sec^{-1}; finally the photoconductive decay time τ microseconds was determined with a pulsed hydrogen–neon time constant apparatus.

The chamber was then evacuated and the layer heated by passing an electric current through it. This heating drove off the adsorbed air, so bringing the layer to a suitably high resistance.

The current through the layer at room temperature, and with 10v across it, was taken as a direct measure of the adsorbed air. Different values of the current through the layer at room temperature were obtained by letting air into the chamber through a needle valve so altering the amount adsorbed in the layer. The measurements listed above were made for a range of currents up to the saturation current at atmospheric pressure. The cycle of measurements was repeated and the results showed that the effect of adsorbed air on the layer was reversible. The results of these experiments for a typical layer are illustrated in the figures. Figure 1 shows the effect on Δi, τ, p and μ of the adsorbed air.

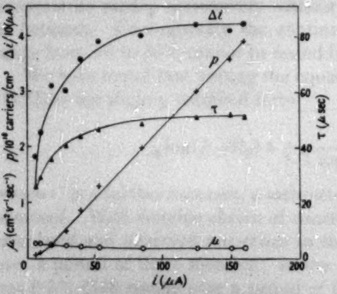


Figure 1. Variation of the time-constant, number of carriers, mobility and photoconductive response of a lead sulphide layer on exposure to air.

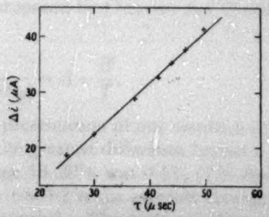


Figure 2. Variation of photoconductive response with time constant as the quantity of adsorbed air varies.

Figure 2 shows Δi plotted against τ using values from the curves in figure 1. It is clear that an increase in Δi is produced by an increase in the amount of adsorbed air up to the saturation value. The number of majority carriers (which are shown by the Hall apparatus to be holes) increase considerably but μ remains relatively constant. As can be seen from figure 2, Δi varies linearly with τ.

Subsidiary experiments in which nitrogen, argon and oxygen were separately admitted to the previously evacuated vessel with no adsorbed gas in the layer have shown that oxygen is the constituent of the air which affects the layer.

These experiments confirm those carried out by Ryvkin and Malakhov (1952) except in one important particular. We find that the number of carriers does vary appreciably with varying amounts of adsorbed air whilst the mobility does not; Ryvkin and Malakhov, however, found that the number of carriers was constant and the mobility varied.

Now the photoconductive equation for Δi can be written

$$\Delta i = A\mathbf{e}[p(\mu-\mu')+2\tau N\mu']$$

where A is a constant, $\mathbf{e}$ is the electronic charge, N is the number of photons absorbed in the layer per second per cubic centimetre and μ' is the mobility under illumination.

From the graph in figure 1 we see that where τ is varying slowly or is substantially constant, Δi behaves in the same fashion; p, however, varies considerably in the same region so that Δi could only remain constant if $\mu-\mu'$ were close to zero. Thus within the experimental error we conclude that the mobility in the layer does not change appreciably on illumination.

It would then appear that a correct interpretation of these results is not a barrier theory as advocated by Ryvkin and Malakhov but a single crystal theory in which the time constant of the layer is increased by an oxygen trap mechanism and the sensitivity of the layer improved.

It is believed that oxygen sensitization of photoconductive layers can in general be accounted for by the increase in effective time constant rather than by the setting up of barriers (Gibson 1951). Experiments are in progress at these laboratories to investigate this hypothesis and it is hoped that a more extended account of this work will be published later.

Mullard Research Laboratories,
Salfords, Redhill, Surrey.
2nd March 1955.

M. Smollett.
R. G. Pratt.

Gibson, A. F., 1951, *Proc. Phys. Soc.* B, **64**, 603.
Ryvkin, S. M., and Malakhov, L. N., 1952, *C. R. Acad. Sci., U.R.S.S.*, **85**, 765.

CONTENTS OF SECTION A

PAGE

Dr. M. J. **Seaton**. Cross Sections for 2s–2p Transitions in H and 3s–3p Transitions in Na Produced by Electron and by Proton Impact 457

Mr. E. D. **Pendlebury**. General Perturbation Theory in Neutronics . . 474

Dr. G. N. **Fowler**. On Star Production by μ-Mesons Underground . . . 482

Dr. L. **Riddiford**, Mr. H. B. **van der Raay** and Mr. R. F. **Coe**. Some Proton Synchrotron Beam Studies with the Induction Electrode 489

Dr. H. **Umezawa**, the late Dr. J. **Podolanski** and Dr. S. **Oneda**. Commutation Relations between Different Fields 503

Mr. J. E. **Bowcock**. Calculation of Reduced Widths from Stripping Curves . 512

Dr. N. L. **Balazs**. Wave Propagation in Even and Odd Dimensional Spaces . 521

Prof. H. **Fröhlich** and Dr H. **Pelzer**. Plasma Oscillations and Energy Loss of Charged Particles in Solids 525

Dr. C. D. **Walshaw**. Line Widths in the 9·6μ Band of Ozone 530

Research Notes:

Dr. H. R. **Allan** and Mr. N. **Sarma**. A Search for Capture Processes in Alpha-Particle and Deuteron Bombardments 535

Mr. A. B. **Clegg**, Dr. G. A. **Jones** and Dr. D. H. **Wilkinson**. The Radiative Capture of Protons of 669 kev in Fluorine 538

Prof. D. R. **Bates** and Dr. B. L. **Moiseiwitsch**. Energies of Normal and Excited Negative Ions 540

Mr. K. **Dressler** and Prof. E. **Miescher**. Spectroscopic Identification of Diatomic PS and New Band Systems of PO and P_2 542

Dr. G. V. **Marr**. The Continuous Atomic Absorption Cross Sections for Neutral Atomic Systems 544

Dr. M. E. **Pillow**. Intensities in Band Systems of CH 547

Letters to the Editor:

Dr. L. R. B. **Elton**. Effect of Finite Nucleus on the Radiative Correction . 549

Mr. E. F. **Bradley**. Energy Loss Distribution for Minimum Ionizing Electrons in a Pressure Proportional Counter 549

Contents of Section B 552

On the Distribution of Impurity in Crystals Grown from Impure Unstirred Melts

By K. F. HULME
H. H. Wills Physical Laboratory, University of Bristol

Communicated by F. C. Frank; MS. received 6th December 1955

Abstract. The existence of a boundary layer of different impurity concentration on the solid–melt interface during crystal growth into impure unstirred melts has been inferred by previous authors. The problem of how rapidly this layer attains its final form in the case of growth into an infinite unstirred melt, hitherto treated approximately, is shown to be capable of exact solution for all values of the segregation constant. The impurity distribution in the resultant solid is calculated and graphs are given for various values of the segregation constant.

---

§ 1. Introduction

When growing crystals by controlled solidification from impure melts it is of interest to know how the impurity will be distributed in the resultant solid. Unless the melt is stirred vigorously (by using an induction furnace to heat it for example) the impurity concentration in the melt is not uniform and a diffusion-controlled boundary layer of different impurity concentration is formed on the solid–melt interface; this boundary layer can under certain circumstances cause 'constitutional supercooling' as described by Rutter and Chalmers (1953); if 'constitutional supercooling' exists the solid–melt interface loses its flatness and the diffusion process ceases to be one-dimensional. The results quoted and derived in this paper are only valid so long as the interface retains its flatness, though they may represent approximations to the truth when the interface becomes corrugated.

Tiller, Jackson, Rutter and Chalmers (1953) have given an expression for the distribution of impurity in the melt a long time after growth has commenced in an infinite melt free from stirring and convection; it is

$$\phi(x)=\phi_0\left[1+\left(\frac{1}{k}-1\right)\exp\left(-\frac{Rx}{D}\right)\right] \qquad \ldots\ldots(1)$$

where $\phi(x)$ is the impurity concentration at a point in the melt at a distance x from the solid–melt interface, ϕ_0 being the constant value $\phi(x)$ tends to as $x\to\infty$, R is the rate of advance of the solid–melt interface with respect to coordinates stationary in the melt, D is the diffusion coefficient of the impurity in the melt and k is the segregation constant of the impurity, i.e. the ratio of the impurity concentration in the solid at the interface to the impurity concentration in the liquid at the interface.

Frank (unpublished work) has considered the effect of taking into account the volume change on solidification. It is then necessary to define concentration more precisely, because the mass of solute per unit volume of solution will not be related to the mass of solute per unit mass of solution by the same factor in the solid as in the liquid. Concentrations in both solid and melt are therefore defined in terms of mass of solute per unit volume and the segration constant k is defined as the ratio of concentrations expressed in these terms.

Then, if there is no change of cross section of the crystal as growth proceeds (e.g. in growth in a vertical cylindrical mould where solidification starts at the bottom), the velocity of the solid–melt interface is different in the two coordinate systems which are respectively at rest in the solid and at rest in the liquid; this is a consequence of the change in volume on solidification. In fact, if R' is the observed velocity in the former system,

$$R'\rho_S = R\rho_L, \quad \text{i.e. } R' = R(1-\alpha)$$

where $\alpha = (\rho_S - \rho_L)/\rho_S$, ρ_L and ρ_S being the densities at the interface of the liquid and solid respectively.

The non-equality of R and R' modifies one of the boundary conditions on $\phi(x)$ (see equation (6) below) and then with $\epsilon = k(1-\alpha)$, so that ϵ is the ratio weight fraction of solute in the solid to weight fraction of solute in the liquid in contact with it, equation (1) becomes, according to Frank,

$$\phi(x) = \phi_0\left[1 + \left(\frac{1}{\epsilon} - 1\right)\exp\left(-\frac{Rx}{D}\right)\right]. \qquad \ldots\ldots(2)$$

From this expression we see that the impurity concentration at the solid–melt interface ($x=0$) is ϕ_0/ϵ; if, when growth commences, the impurity concentration is uniform it will be ϕ_0. The problem of how rapidly the transition from the concentration $\phi = \phi_0$ to $\phi = \phi_0/\epsilon$ takes place at $x=0$ has been treated by Tiller *et al.* (1953) and by Pohl (1954); the solutions obtained are not exact but are good approximations for small values of k. The analysis given below is closely similar to that used by Pohl (1954); it differs in the treatment of one of the poles of the integrand (at $p = -4\epsilon(1-\epsilon)$) involved in the inversion of the Laplace transform and in that the integrals are evaluated exactly. The resulting solution does not give anomalous behaviour for large values of ϵ as the previous solutions did.

§ 2. THE DIFFERENTIAL EQUATION AND BOUNDARY CONDITIONS

In the analysis given below it will be assumed that (1) R, D and k are constants; assuming the constancy of k implies idealization of the actual phase diagram and confining our attention to cases in which the solid will contain only one phase. (2) The melt is completely free from stirring or convection currents. (3) The cross-sectional area is constant. (4) The melt is infinite. (5) The interface remains flat.

The axes of reference are chosen so that the axis of x is along the direction of growth and normal to the solid–melt interface at $x=0$; the melt then fills space for $x>0$. If $\phi(x, t)$ is the impurity concentration per unit volume assumed independent of (y, z), the differential equation obeyed by ϕ is (Wilson 1904)

$$D\frac{\partial^2\phi}{\partial x^2} + R\frac{\partial\phi}{\partial x} = \frac{\partial\phi}{\partial t}. \qquad \ldots\ldots(3)$$

The boundary conditions to be satisfied by ϕ will now be considered. If the impurity concentration in the melt is uniform at the start then

$$\phi = \phi_0 \text{ at } t=0 \text{ for all } x. \qquad \ldots\ldots(4)$$

Further, since the impurity concentration at large x will be only slightly affected by the conditions near $x=0$, it may be inferred that

$$\phi \to \phi_0 \text{ as } x \to \infty \text{ for all } t. \qquad \ldots\ldots(5)$$

Finally there is a boundary condition at the plane $x=0$ expressing the fact that there is conservation of impurity. Consider the fluxes of impurity in and out of the region between the planes $x=0$, $x=\delta x$; then from figure 1:

$$\delta x \left.\frac{\partial \phi}{\partial t}\right|_{x=x_1} = R\phi(\delta x, t) - R(1-\alpha)k\phi(0,t) + D\left.\frac{\partial \phi}{\partial x}\right|_{x=\delta x}$$

where x_1 is some value of x in the range 0, δx.

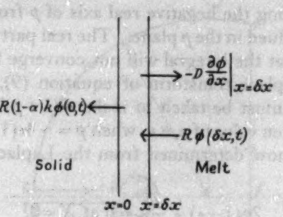


Figure 1.

As $\delta x \to 0$ this gives the condition

$$R\phi[1-k(1-\alpha)] + D\frac{\partial \phi}{\partial x} = 0 \text{ at } x=0 \text{ for all } t. \qquad \ldots\ldots(6)$$

Putting $k(1-\alpha)=\epsilon$ and introducing the dimensionless variables $X=Rx/2D$, $T=R^2t/4D$ the differential equation (3) becomes

$$\frac{\partial^2 \phi}{\partial X^2} + 2\frac{\partial \phi}{\partial X} = \frac{\partial \phi}{\partial T}. \qquad \ldots\ldots(7)$$

The boundary conditions (4), (5), (6) are now:

$$\phi = \phi_0 \text{ at } T=0 \text{ for all } X \qquad \ldots\ldots(8)$$

$$\phi \to \phi_0 \text{ as } X \to \infty \text{ for all } T \qquad \ldots\ldots(9)$$

$$2\phi(1-\epsilon) + \frac{\partial \phi}{\partial X} = 0 \text{ at } X=0 \text{ for all } T. \qquad \ldots\ldots(10)$$

§ 3. SOLUTION OF THE EQUATIONS

The Laplace transform is now applied to equation (7). If the Laplace transform of ϕ (X, T) is $\bar{\phi}(X, p)$, i.e.

$$\bar{\phi}(X,p) = \int_0^\infty e^{-pT}\phi(X,T)\,dT$$

the Laplace transforms of $\partial^2\phi/\partial X^2$ and $\partial\phi/\partial X$ are respectively $d^2\bar{\phi}/dX^2$, $d\bar{\phi}/dX$, since integration with respect to t and differentiation with respect to X are commuting operations on ϕ in this case. But the Laplace transform of $\partial\phi/\partial T$ is

$$\int_0^\infty e^{-pT}\frac{\partial \phi}{\partial T}\,dT = \Big[e^{-pT}\phi\Big]_0^\infty + p\int_0^\infty e^{-pT}\phi\,dT = -\phi_0 + p\bar{\phi}.$$

Equation (7) therefore transforms to

$$\frac{d^2\bar{\phi}}{dX^2}+2\frac{d\bar{\phi}}{dX}-p\bar{\phi}=-\phi_0. \qquad \ldots\ldots(11)$$

This has the general solution

$$\bar{\phi}=\frac{\phi_0}{p}+e^{-X}(A\exp[X(1+p)^{1/2}]+B\exp[-X(1+p)^{1/2}]). \qquad \ldots\ldots(12)$$

The solution ϕ will be determined from the formula

$$\phi=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\bar{\phi}e^{pT}dp. \qquad (\gamma>0).$$

A cut is inserted along the negative real axis of p from $p=-1$ to $p=-\infty$ to make $(1+p)^{1/2}$ single-valued in the p plane. The real part of $(1+p)^{1/2}$ will now tend to ∞ as $\mathscr{I}(p)\to\infty$ so that the integral will not converge for any X greater than 0 unless $A=0$. The Laplace transform of equation (9), $\bar{\phi}\to\phi_0/p$ as $X\to\infty$, also requires $A=0$. (Care must be taken to make $\mathscr{R}(1+p)^{1/2}\geqslant 0$ when the value of $(1+p)^{1/2}$ is used for a given value of p, e.g. when $p=-4\epsilon(1-\epsilon)$, $(1+p)^{1/2}=|1-2\epsilon|$).

The value of B is now determined from the Laplace transform of equation (10), i.e.

$$2\bar{\phi}(1-\epsilon)+\frac{d\bar{\phi}}{dX}=0 \text{ at } X=0.$$

Thus $B=\dfrac{2(1-\epsilon)\phi_0}{p[(1+p)^{1/2}-(1-2\epsilon)]}$.

The solution of equation (11) appropriate to this problem is therefore

$$\bar{\phi}=\frac{\phi_0}{p}+\frac{2(1-\epsilon)\phi_0\exp\{-X[1+(1+p)^{1/2}]\}}{p[(1+p)^{1/2}-(1-2\epsilon)]}.$$

The inversion formula for the Laplace transform is now used to obtain $\phi(X,T)$ from $\bar{\phi}(X,p)$. Then, since the inverse of ϕ_0/p is simply ϕ_0,

$$\phi=\phi_0+\frac{1}{2\pi i}2(1-\epsilon)\phi_0 e^{-X}\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{\exp(pT)\exp[-X(1+p)^{1/2}]\,dp}{p[(1+p)^{1/2}-(1-2\epsilon)]} \qquad \ldots\ldots(13)$$

where γ is greater than the real parts of all the singularities of the integrand.

There is a simple pole of the integrand at $p=0$; the residue R_0 is given by

$$R_0=\lim_{p\to 0}\left(p\frac{\exp(pT)\exp[-X(1+p)^{1/2}]}{p[(1+p)^{1/2}-(1-2\epsilon)]}\right)=\frac{e^{-X}}{2\epsilon}. \qquad \ldots\ldots(14)$$

The integrand also has a singularity where

$$(1-2\epsilon)=(1+p)^{1/2}, \quad \text{i.e. } p=-4\epsilon(1-\epsilon)=p_1$$

provided $(1-2\epsilon)\geqslant 0$, i.e. $\epsilon\leqslant\frac{1}{2}$ (because with $p=-4\epsilon(1-\epsilon)$, $(1+p)^{1/2}=|1-2\epsilon|$ and this is only equal to $(1-2\epsilon)$ if $\epsilon\leqslant\frac{1}{2}$). This singularity is therefore ignored unless $\epsilon\leqslant\frac{1}{2}$.

The behaviour at this singularity must be examined. Putting $p=p_1+Z$

$$(1+p)^{1/2}=(1-4\epsilon+4\epsilon^2+Z)^{1/2}=(1-2\epsilon)[1+Z/(1-2\epsilon)^2]^{1/2}.$$

Expanding by Taylor's theorem about $Z=0$

$$(1+p)^{1/2}-(1-2\epsilon)=\frac{Z}{2(1-2\epsilon)}(1+\ldots).$$

It therefore appears that the residue at p_1 is

$$R_{p_1} = -\frac{(1-2\epsilon)\exp(p_1 T)\exp[-X(1-2\epsilon)]}{2\epsilon(1-\epsilon)}. \qquad \ldots\ldots(15)$$

Further, it is noted that the integrand has a branch point at $p=-1$.

The integral

$$I_0 = \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{\exp[pT - X(1+p)^{1/2}]\,dp}{p[(1+p)^{1/2}-(1-2\epsilon)]}$$

is now evaluated by contour integration in the complex p plane. The contour used is shown in figure 2.

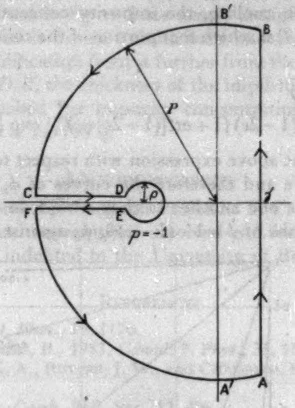


Figure 2.

The contour must not cross the axis of real p for p less than -1 or a different branch of the integrand is entered; hence the deformation of the contour to encircle the branch point.

It may be shown that

$$\int_{BB'} \to 0, \quad \int_{B'C} \to 0, \quad \int_{FA'} \to 0, \quad \int_{A'A} \to 0 \text{ as } P \to \infty$$

and that

$$\int_{DE} \to 0 \text{ as } \rho \to 0.$$

As $P \to \infty$ it is seen that $\int_{AB} \to I_0$.

Using Cauchy's residue theorem, we have as $P \to \infty$, $\rho \to 0$,

$$I_0 = 2\pi i(R_0 + R_{p_1}') - \int_{CD} - \int_{EF} \qquad \ldots\ldots(16)$$

where $R_{p_1}' = R_{p_1}$ for $\epsilon < \frac{1}{2}$, $R_{p_1}' = 0$ for $\epsilon \geqslant \frac{1}{2}$.

Putting $p = -1 + Se^{i\pi}$ along CD and $p = -1 + Se^{-i\pi}$ along EF and letting $P \to \infty$ the integrals along CD and EF may be combined. The resulting integral

$$\int_{CD} + \int_{EF} = 2ie^{-T} \int_0^\infty \frac{(\sqrt{S}\cos X\sqrt{S} - (1-2\epsilon)\sin X\sqrt{S})\exp(-ST)\,dS}{(S+1)[S+(1-2\epsilon)^2]}$$

can be evaluated for general X to give $\phi\,(X, T)$ exactly, but since the main interest is in the distribution of the solute in the solid and this depends only on $\phi\,(0, T)$ it has only been evaluated here for $X=0$. The results are different for $\epsilon<\frac{1}{2}$ and for $\epsilon\geqslant\frac{1}{2}$; the expression for I_0 is, however, the same for all values of ϵ because of the peculiarity of the residue at $p=p_1$.

Using these results and equations (13), (14), (15), (16) with $X=0$ we obtain, for all values of ϵ,

$$\phi(0, T)=\frac{\phi_0}{2\epsilon}\,[(1+\operatorname{erf}T^{1/2})-(1-2\epsilon)\{1+\operatorname{erf}[(1-2\epsilon)T^{1/2}]\}\exp[-4\epsilon(1-\epsilon)T].$$

If there is no diffusion of impurity in the solid then, ignoring the changes of length of the crystal on cooling, the impurity concentration in the solid as a function of the value of T at which that portion of the solid was formed is given by

$$\phi_S=k\phi(0, T)=\frac{\phi_0}{2(1-\alpha)}$$
$$\times\Big[(1+\operatorname{erf}\sqrt{T})-(1-2\epsilon)\{1+\operatorname{erf}[(1-2\epsilon)\sqrt{T}]\}\exp\,[-4\epsilon(1-\epsilon)T]\Big]\,.\;.\,(17)$$

By differentiating the above expression with respect to ϵ it may be shown that $\partial\phi_S/\partial\epsilon$ is always positive and therefore the curves of ϕ_S against T for different values of ϵ do not cross one another as they would do for Pohl's solution. In figure 3 are shown graphs of $\psi=1-(1-\alpha)\phi_S/\phi_0$ against T for various values of ϵ; ψ is then a measure of the departure of ϕ_S/ϕ_0 from the value $1/(1-\alpha)$ that it approaches as $T\to\infty$. The graphs may be compared with similar ones published by Pohl (1954). In using these graphs we remember that since the growth velocity is constant a particular value of T corresponds to a particular plane in the resultant crystal and that T is therefore a measure of distance from the plane at

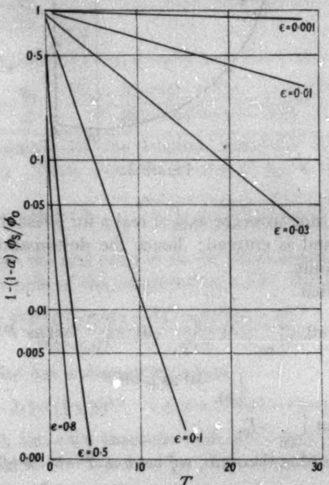


Figure 3.

which solidification is assumed to have started at $T=0$. For $R=10^{-3}$ cm sec^{-1} ($=3{\cdot}6$ cm h^{-1}) and $D=2{\cdot}5\times10^{-5}$ cm^2 sec^{-1}, $T=10$ gives this distance $4DT/R$ as 1 cm.

For $(1-2\epsilon)\,T^{1/2}>2$ and $\epsilon<\frac{1}{2}$ a useful approximation is

$$\psi\sim(1-2\epsilon)\exp[-4\epsilon(1-\epsilon)T];$$

for then erf $T^{1/2}$ and erf $(1-2\epsilon)T^{1/2}$ are very close to unity. A further approximation is $\psi\sim\exp[-4\epsilon(1-\epsilon)T]$. The value of $\epsilon(1-\epsilon)T$ for ψ to fall to 0·01 obtained from the latter approximation is 1·15 compared with the values obtained by using equation (17) of 1·15, 1·15, 1·13, 1·09, 0·70 for $\epsilon=0{\cdot}001$, 0·01, 0·03, 0·1, 0·5 respectively. The closeness of the approximation for low ϵ accounts for the fact that many of the curves in figure 3 are nearly straight lines.

The solution obtained is for an infinite melt and should be applicable to a finite melt when the solidification front is further from the termination of the melt than a few multiples of D/R, the thickness of the impurity boundary layer; after this stage has been reached the impurity concentration will rise rapidly (see qualitative graph, Tiller *et al.* 1953).

ACKNOWLEDGMENTS

I would like to express my thanks to Professor F. C. Frank for discussion of the fundamental facts and assumptions involved and for much helpful criticism and advice. I am also indebted to the University of Bristol for a maintenance grant.

REFERENCES

POHL, R. G., 1954, *J. Appl. Phys.*, **25,** 1170.

RUTTER, J. W., and CHALMERS, B., 1953, *Canad. J. Phys.*, **31,** 15.

TILLER, W. A., JACKSON, K. A., RUTTER, J. W., and CHALMERS, B., 1953, *Acta Metallurgica*, **1,** 428.

WILSON, H. A., 1904, *Proc. Camb. Phil. Soc.*, **12,** 406.

Metal Transfer and the Wear Process

By M. KERRIDGE

Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berkshire

Communicated by W. Hirst; *MS. received 24th January* 1955

Abstract. Experiments have been made with a radioactive, annealed steel pin rubbing against a hardened steel ring to compare the amount of wear with the amount of metal transferred from one surface to the other by welding. It was found that a layer of transferred matter from the pin was formed on the surface of the ring and that an equilibrium condition was reached in which the loss of wear debris from this layer was balanced by further transfer. By using a combination of radioactive and inactive test-pieces the rate of transfer to the ring in the equilibrium condition could be estimated and was found to be the same as the wear rate of the pin. There was no evidence that any loose wear particles were formed by direct removal of material from the pin. The transferred layer became oxidized due to frictional heating and it is concluded that the wear process consisted of three stages: transfer of metal, oxidation of the transferred layer, and the subsequent removal of the oxide to form a loose wear-product.

§ 1. INTRODUCTION

It is generally accepted that the formation and rupture of welded junctions between the rubbing surfaces is an important factor in the wear of metals (Bowden and Tabor 1950). If the rupture occurs away from the original interface, the result is that particles of metal from one surface remain adhering to the other. This process of ' metal transfer ' can be studied conveniently by making one of the rubbing members radioactive and detecting the transfer to the inactive surface by radioactivity measurements. A number of workers have applied such techniques in investigations of transfer between metal surfaces sliding together at low speeds (Gregory 1946, Sakmann, Burwell and Irvine 1944, Rabinowicz and Tabor 1951, Kerridge 1952, Rabinowicz 1953). The experiments have shown that localized welding takes place both for dry and for lubricated metals and have yielded information about the process of metal transfer. On the other hand, there has been little investigation of the relation between metal transfer and metallic wear, by which is usually meant the removal of material from the rubbing members to form a loose wear-product.

As Feng (1952) has pointed out, intermetallic welding does not, by itself, produce a loose wear-product ; it is therefore an open question whether the wear particles leaving the system are formed directly by some process other than welding, abrasion for instance, or whether welding is one of a succession of steps leading to the formation of the loose particle. In the theories of wear which have recently been proposed to explain the laws observed under certain simplified conditions (Burwell and Strang 1952, Archard 1953) the postulate is made that the number of wear particles produced is proportional to the number of encounters between the asperities on the two surfaces. By assuming that the proportionality factor or 'wear coefficient' is constant the correct dependence of the amount of wear upon the sliding distance and the applied load can be derived; however, the magnitude of the wear coefficient cannot be accounted for without more knowledge of the mechanism of wear than now exists. In the present investigation it is shown that the combination of radioactive tracer methods with the more usual type of wear measurement, thereby permitting the study of wear and transfer together, offers a means of elucidating certain aspects of the wear mechanism. It appears that the process leading to the formation of a loose wear particle may comprise several stages. Although the method is of wide applicability, the particular wear mechanism described below is offered as applying only to the materials examined and the experimental conditions used.

§ 2. APPARATUS AND MATERIALS

The machine used for the test was a simple pin and ring apparatus similar to that employed by Archard (1953). In this equipment a flat-ended pin is loaded against the cylindrical surface of a rotating ring and the wear of the pin is determined from the dimensions of the wear scar produced on it. In the experiments the ring was made harder than the pin and its wear was negligible compared with that of the latter. The pins were made radioactive by pile irradiation at Harwell so that the amount of metal transferred to a ring could be measured by means of a calibrated, end-window Geiger–Müller counter. On the machine the pin was held in a loading arm which pivoted on accurate ball bearings and the apparatus was arranged so that the arm could be taken off and replaced without upsetting the alignment of the test-pieces. In an experiment the procedure was to run the pin and ring together for the desired time and then remove the arm in order to measure the wear scar of the pin under a microscope and to allow the amount of active metal transferred to the ring to be determined without abnormal background interference. The counter was mounted in a block and was within two or three millimetres of the ring surface during the activity measurements which were made with the ring rotating. At other times it was swung clear of the specimens to prevent it becoming contaminated with active wear powder.

The rings were of hardened tool steel (Vickers D.P.N.860), their diameter was 2·4 cm and the rubbing surface had a ground finish (Talysurf C.L.A. value about 6 microinch). They were rotated at 1450 rev min^{-1} to give a rubbing speed of about 180 cm sec^{-1}. The pins had flat sides parallel to the sliding direction and were 5·4 mm in width thus giving initially a nominal line contact of this length with the ring. Experiments were made with pins of three types of steel but, since the behaviour was similar in each case, only the results with annealed tool steel (V.D.P.N.270) will be discussed. This had a composition: 0·7% C, 18% W, 4·5 Cr, 1·5% V; the other steels were a 3% chromium–molybdenum steel (B.S. No. En.29), and a 17% chromium, corrosion-resisting steel.

§ 3. RESULTS

3.1. *Amount of Transfer*

Using various loads in the range 150–1500 g it was found, with fresh test-pieces, that the amount of transferred material on a ring increased during a run until it reached a limiting value at which it remained constant. This limiting amount of transfer was independent of the magnitude of the load although the exact value was rather variable from test to test being between 0·7 and 1·0 mg. The time taken to reach the steady condition was, however, related to the load; it was roughly 10 minutes and 100 minutes at the heaviest and lightest loads respectively.

The wear of the pin was measured at intervals during each run and was found to increase linearly with time. The rates of wear ranged from 33 μg min^{-1} at 150 g load to 160 μg min^{-1} at 1500 g. The initial transfer rates were estimated to be somewhat larger than the average rates of wear throughout the runs. Typical results are shown in figures 1 and 2 but, because of the limited sensitivity of the wear scar measurements, the relatively high initial wear rate which is seen on the graph for 1500 g load could not, in general, be detected.

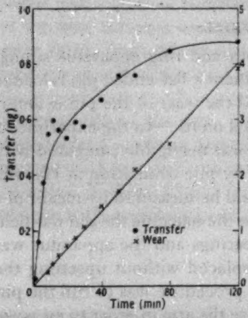


Figure 1. Wear and transfer at a load of 150 g.

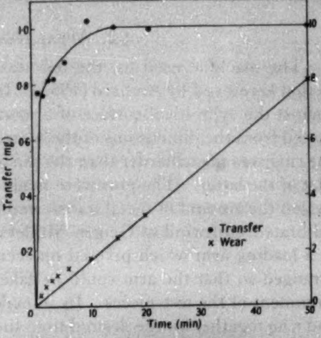


Figure 2. Wear and transfer at a load of 1500 g.

3.2. *Behaviour of the Transferred Material*

A further series of experiments was carried out to discover if the transfer was incidental to the wear process or played an essential part in it. In these tests the aim was to study whether or not transfer continued despite the limit to the amount of transferred material remaining on the ring. The procedure was to run a radioactive pin against a ring for a time considerably longer than that required to establish the limiting amount of transferred material. A similar run was then made for the same time with another pair of test-pieces but using an inactive pin. The ring prepared in the first run was then substituted for the second ring and the run continued for the same time again. Precautions were taken to ensure identical alignment of the test-pieces; these were apparently satisfactory since no discontinuity was evident in the wear–time curves which also were the same as in control experiments where no change of rings was made. It was found that, as soon as the change-over was effected, the amount of activity on the ring began to decrease and gradually fell virtually to zero.

These results show that, although in experiments of the type illustrated by figures 1 and 2 the amount of active material from the pin which is on the ring remains constant, the transferred layer is nevertheless continually being removed and must play some part in the wear process. An attempt was made next to determine the quantitative relationship between the equilibrium rate of replacement of transferred material and the equilibrium wear rate. In the preceding experiments the active pins had been run against freshly prepared rings and it can be seen from figures 1 and 2 that the initial rate of transfer to the ring (and therefore the initial rate of loss of material from the pin) exceeded the average wear rate as determined by the rate of increase in size of the wear scar on the pin. To determine the equilibrium relationships the experiments were repeated in the reverse order, first running an inactive pin against the ring until an inactive, equilibrium, transferred layer had been developed. Then, by changing to an active pin, the continuance of transfer and eventual replacement of the original material was observed as a growth instead of a fall in the activity of the ring.

Repeat tests were also made in which the inactive layers had been produced by successive runs with active and then inactive pins. The limiting activity of the transferred layer was the same as its original value ; this confirmed that the quantity of transferred material remained the same throughout, independent of the changes of pin. It was also found that the behaviour was the same when the substitution pins were unworn as when they had already been run against a ring to develop an appreciable wear scar. This was as expected, since, on this type of machine, the wear rate is independent of the apparent area of contact which continually increases during a run.

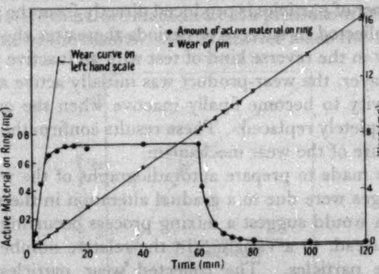


Figure 3. Conversion of an equilibrium, inactive transferred layer to an active layer and vice versa after changing between inactive and active pins. Load 1000 g.

In the various experiments the equilibrium transfer rates, as far as could be estimated, were the same as the steady wear rates. Figure 3 shows the results of an experiment of this kind at a load of 1000 g. An equilibrium, inactive layer had been formed previously on the ring and the origin of the graph corresponds to the instant when the inactive pin was replaced by an active pin. After 60 minutes the active pin was itself replaced by an inactive one. The figure shows that the pin wear rate remains constant throughout and is indentical, within experimental error, with the initial rate of increase of the amount of active material, i.e. the equilibrium transfer rate.

The agreement between the magnitudes of the rate of wear and the rate of replacement of the transferred material strongly suggests that the wear particles are produced by an indirect process, one stage consisting of entry into the transferred layer. Moreover, the fact that on a fresh ring the initial rate of transfer is greater than the equilibrium value indicates that the factor determining the wear rate is the rate of removal of material from the transferred layer not the rate of formation of the layer.

3.3. *Absence of Back Transfer*

In the tests in which inactive pins were run on an active transferred layer the wearing surface of the pin was examined for activity. None was detected and therefore the wear particles must be formed from the layer on the ring without any further stage involving back transfer to the pin.

3.4. *Examination of the Wear Product*

Although the apparent identity of the equilibrium wear and transfer rates, taken together with the absence of back transfer, indicates that the wear particles are formed from the transferred layer, there yet remains the possibility that a proportion of the particles (too small to make a detectable difference between the wear and transfer rates) are removed directly from the pin. Tests were made to examine this question by observing the activity of the wear particles during the runs in which a change between active and inactive pins was made. When an active pin replaced an inactive one, it was found that the wear particles produced immediately afterwards were inactive ; therefore no detectable number of particles is produced directly from the pin. The samples of wear-product collected in successive periods thereafter showed an increasing activity. Similarly in the reverse kind of test with an inactive pin rubbing on an active transferred layer, the wear-product was initially active and then gradually diminished in activity to become finally inactive when the original transferred layer had been completely replaced. These results confirm the earlier hypothesis concerning the nature of the wear mechanism.

An attempt was made to prepare autoradiographs of the wear powder to see if the activity changes were due to a gradual alteration in the specific activity of the material (which would suggest a mixing process occurring within the transferred layer) or, instead, to a variation in the relative numbers of individually active and inactive particles. The collected wear particles were, however, aggregates of smaller particles and the results were inconclusive.

§ 4. MECHANISM OF REMOVAL OF THE WEAR PARTICLES

The wear detritus consisted of a fine, brown-coloured powder ; x-ray powder photographs showed that it was either hematite ($\alpha-Fe_2O_3$) or hydrohematite ($\alpha-Fe_2O_3 . nH_2O$) the diffraction patterns of which are similar. A comparison was made with a glancing-angle, x-ray photograph of a pin and it was found that the wear powder gave no lines corresponding to the pin material except for a very faint line due to the strongest reflection from α-iron.

The rings were examined under the microscope and it was observed that the wear track was covered to a large extent with a smooth, thin layer of grey material on which were patches of finely divided, brown, oxide powder. This powder could easily be removed by gentle rubbing whereas the grey layer was firmly attached. The original surface of the ring could be seen in some places and it appeared to be undamaged. Using the interferometer microscope described by Dyson (1953), it was confirmed that the grey layer was raised and therefore constituted the actual rubbing surface. Its thickness was estimated as between one and two fringes (1 fringe = 0·27 micron) and this is consistent with the amount of transfer determined by the radioactivity method ; one milligram of material spread uniformly over the wear track would give a thickness of 0·28 micron.

Autoradiographs of rings which had run against active pins were prepared and showed an exact correspondence between the activity and the raised layer.

There can therefore be little doubt that the grey layer was derived entirely from the material of the pin. Micro-hardness tests indicated that it was considerably harder than the original pin material and this may be due to the intense, localized, thermal effects which, as is well known, accompany rubbing under unlubricated conditions. The increased hardness of the transferred over the original material is probably the cause preventing back transfer from ring to pin and makes it unlikely that the wear particles are formed by removal of the transferred, grey material itself. Of the two possibilities, that the oxidized wear particles are formed by oxidation before or after removal from the ring, the former would appear the more probable, particularly in view of the presence of loosely adhering, brown powder on the ring surface.

If these views are correct, and if, as suggested earlier, it is true that the wear rate is not controlled by the rate of transfer of material from pin to ring, the factor determining the wear rate would be the rate of oxidation of the transferred layer. Thus a reduction in the rate of oxidation should lead to a diminished wear rate. Wear experiments were therefore made in air at a pressure of 10^{-3} mm of mercury ; figure 4 shows some of the results. At the heavier loads the wear rates were less than one tenth of the wear rates in air at atmospheric pressure and at the lighter loads the reduction was even greater. Patches of the grey material were formed on the rings but only very small amounts of the brown powder could be discerned. The wear particles collected were oxide.

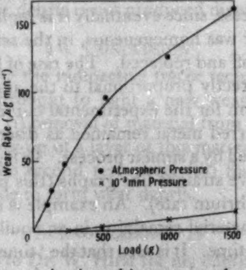


Figure 4. Wear rates in air and in vacuum plotted against load.

§ 5. DISCUSSION

It is clear from the experimental results that the wear mechanism, under the particular conditions, is a three-stage process involving transfer of metal, its conversion to oxide and the removal of oxide to give a loose wear-product. It is suggested that the behaviour is as follows. When the test-pieces first rub together, localized welding takes place and metal is transferred from the soft pin to the harder ring. At this stage the rate of transfer is greater than the wear rate which is found under steady conditions. The transfer occurs at the actual regions of contact and, since these are continually changing, a layer of transferred material gradually forms over the whole of the wear track on the ring. The real area of contact is proportional to the load and hence, as is found, there will be an inverse relationship between the load and the time taken to produce a complete layer.

The layer of transferred material is apparently formed by a smearing-out and hardening of individual, welded fragments. Since no back transfer from ring to pin was detected, it seems that the layer itself does not readily form welds (this view is supported by the reduction in wear found in the vacuum experiments). The high temperatures developed at the actual contact regions cause the transferred material to become oxidized and the oxide, not being strongly adherent, is rubbed off by the repeated sliding. At equilibrium the transferred layer is being removed in this way at a steady rate and the wear of the pin arises from welding and transfer to regions of the original ring surface as these become exposed. A balance thus exists between the transfer rate and the rate at which a loose wear-product is formed.

In the equilibrium experiments, therefore, when an active pin is substituted for an inactive one, the inactive layer is gradually replaced by active material. The initial rate of increase of the amount of active material is a measure of the equilibrium transfer rate since, at this stage, activity is being transferred into the layer while only inactive material is being worn off. As the proportion of active material in the equilibrium layer increases the proportion worn off rises similarly, the rate of conversion of an inactive to an active layer therefore lessens.

The experiments also demonstrate that the whole of the equilibrium layer takes part in the wear process, since eventually it is replaced completely. It might be assumed that the layer was homogeneous, in the sense that all parts of it were equally likely to be worn off and replaced. The rate of loss of active material from a layer would then be directly proportional to the amount in it and this would lead to an exponential form for the experimental curves. The same relationship would arise if the transferred metal remained as discrete fragments on the ring surface, each being removed by a similar process. In fact the results, when plotted logarithmically, do not give straight line graphs (this precludes the use of such a plot to estimate the equilibrium rate). An example is shown at figure 5 in which the percentage of active material remaining in an equilibrium layer is plotted on a logarithmic scale against time. It is seen that the 'time constant' for replacement gradually increases. This may be interpreted as due to a variation in the probability of oxidation and removal for different portions of the layer and is not inconsistent with the hypothesis that the original transferred fragments are smeared out; in such a case a preferential removal of the more recently transferred material, at the expense of the underlying parts of the layer, would be expected.

The addition of the tracer technique to the more usual methods of investigating wear has, therefore, made it possible to obtain quite a detailed picture of the stages of the wear process. Without the evidence of the radioactive experiments the interpretation of the remaining results would have been far more difficult. Thus, referrring to the reduction in the wear rate at low air pressure, several alternative explanations would have been possible. Fink and Hofmann (1932) observed that, under certain conditions, wear rates could be reduced by running in an inert atmosphere and supposed that in air the surfaces are chemically activated by running together and form oxide which, by penetrating and destroying the coherence of the metal, causes wear particles to become detached. Mailander and Dies (1943) have also observed that running at a low air pressure may considerably reduce the wear of steels and Dies (1943) has concluded that if hard oxides form on a rubbing surface they may abrade the opposite member. Neither of these processes can apply to the present series of experiments since the tracer method shows that the wear debris is not produced directly from the pin.

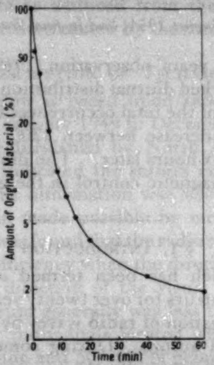


Figure 5. Proportion of active material remaining in an equilibrium layer plotted on a logarithmic scale against time after replacement of active pin by an inactive pin.

§ 6. Conclusion

The study of wear using the radioactive tracer technique combined with the more usual methods has brought to light a new wear mechanism. The wear process has several steps : (i) removal of material from one test-piece by transfer to the other, (ii) the building-up of a layer of this material, but as a different hard form, on the surface of the second test-piece, (iii) the oxidation of this layer, (iv) the removal of the oxide to form wear debris.

Acknowledgments

The author is indebted to Dr. W. Hirst for his continued interest and advice ; he also thanks Dr. T. E. Allibone for permission to publish this paper.

References

Archard, J. F., 1953, *J. Appl. Phys.*, **24**, 981.
Bowden, F. P., and Tabor, D., 1950, *The Friction and Lubrication of Solids* (Oxford : Clarendon Press).
Burwell, J. T., and Strang, C. D., 1952, *J. Appl. Phys.*, **23**, 18.
Dies, K., 1943, *Arch. Eisenhüttenw.*, **16**, 399.
Dyson, J., 1953, *Proc. Roy. Soc.* A, **216**, 493.
Feng, I.-Ming., 1952, *J. Appl. Phys.*, **23**, 1011.
Fink, M., and Hofmann, U., 1932, *Arch. Eisenhüttenw.*, **6**, 161.
Gregory, J. N., 1946, *Nature, Lond.*, **157**, 443.
Kerridge, M., 1952, *Radioisotope Techniques*, **2**, 26 (London : H.M.S.O.).
Mailander, R., and Dies, K., 1943, *Arch. Eisenhüttenw.*, **16**, 385.
Rabinowicz, E., 1953, *Proc. Phys. Soc.* B, **66**, 929.
Rabinowicz, E., and Tabor, D., 1951, *Proc. Roy. Soc.* A, **208**, 455.
Sakmann, B. W., Burwell, J. T., and Irvine, J. W., 1944, *J. Appl. Phys.*, **15**, 459.

Diurnal Variation in Auroral Activity

By N. C. GERSON
Geophysics Research Directorate, Air Force Cambridge Research Center,
Air Research and Development Command

MS. received 16*th August* 1954, *and in final form* 7*th February* 1955

Abstract. Analysis of three years' observation (1949–1951) of the ionized aurora indicate that it displays a marked diurnal distribution having a strong maximum at 2100 L.S.T. and having 86% of the total occurrences between 1700 and 2400 L.S.T. Auroral activity begins to increase between 1200 and 1500 L.S.T., attaining a nocturnal maximum about six hours later. The diurnal variation does not seem to be dependent upon geomagnetic control in the geomagnetic latitude interval 50–60° N.

---

§ 1. INTRODUCTION

THE phenomenon which has been termed *auroral interaction* has been known to the radio amateurs for over twenty years (Wilson 1941). Grossly, it describes the propagation of radio waves by means of refraction from or interaction with the ionized aurora. During some periods when the ionized aurora changes its characteristics rapidly, the net result is similar to that described by the Luxembourg effect or radio-wave interaction. In the latter case, the electron motion which modulates the wanted signal arises from a man-made source—the transmitter producing the unwanted signal. In the former instance, however, the often observed modulation imposed on the wanted signal arises from electron motions associated with the incoming auroral primaries (or possibly secondaries). Although the Luxembourg effect defines an intelligible modulation, auroral interaction defines a random, statistical 'noise' modulation.

Auroral interaction, of course, may be minor, moderate or severe. When minor, the modulation may be unnoticeable; the signals returned from the ionized aurora are steady and clear as though reflected from a stable ionic layer. Contrarily, where auroral interaction is severe, the returned signals are unintelligible; the signals show severe garbling, and 'hissing' is common; radiotelephone cannot be employed; continuous wave radiations are necessary and even then may not be completely identified. In the extreme, when auroral interaction is severe, reflections from the ionized aurora may be worse than useless: not only do they not allow identification of the desired signal but they increase the general noise level at the frequency employed.

Observations of auroral interaction have been made by radio amateurs operating in the frequency range of 30–150 Mc/s and have been noted as interference to television and ultra-high frequency signals to a frequency of 200 Mc/s. During the present sunspot minimum, signals returned from the aurora have frequently been very steady, at times indicating a minimum of modulation. Some increase in this noise modulation would be expected during sunspot maximum.

Investigations of auroral interaction have proceeded along two lines: (*a*) studies by means of radio wave probings at essentially normal incidence to the ionized aurora (Aspinall and Hawkins 1950, Forsyth 1954, Forsyth, Petrie and Currie 1949, Harang and Landmark 1954, Harang and Stoffregen 1940, Lovell, Clegg and Ellyett 1947, McKinley and Millman 1953, 1954, Peterson and Leadabrand 1954); and (*b*) analyses of the radio contacts established between two stations by radio waves incident obliquely to the ionized aurora (Ferrell 1948, Gerson 1951 a, b, Moore 1951, Wilson 1941). This paper describes the diurnal variation in the ionized aurora as deduced from the latter type of study and compares the results with those determined by the former method.

§ 2. Description of Technique

The analysis presented below is based upon reports of auroral interaction which occurred in the interval January 1949 to December 1951 inclusive (Gerson 1951 b). These reports were furnished by North American radio amateurs. Their voluntary observations indicated the station contacted, and the time the radio contact was made. Some information was also provided on the duration of the two-way communication made possible by means of reflection from the aurora. In almost all cases, it was found that the radio contact could be established and maintained with greatest efficiency when the directive antennae were pointed northward somewhere in the direction of the visual aurora. The transmitted power used by the radio amateurs generally was less than 80 watts. Practically all the data were obtained by individuals located within the area bounded by 70°–125° W geographic longitude and 50°–60° N geomagnetic latitude.

In preparing the summary, all available information furnished by the amateurs was lumped together. A common standard of 75° W meridian time was used. The data were tabulated by hours of the day for each month of the year. In this connection, a given hour (XX) was considered to include the interval from (XX – 1)30 to (XX)29 inclusive. The occurrence of auroral propagation any time in this hourly period was considered as *one case*, when reported by three or more pairs of observers. The total number of aurorae reported was 135 which occurred on 149 calendar days.

The total number of hours during which communication was maintained by means of the aurora was about 440. While this number is not necessarily equal to the total number of hours during which communication was possible, it does provide a fair sample suitable for some general conclusions. When tabulated, these data constituted 581 cases. No differentiation was made between strong or weak signals, or between mild, moderate or severe modulation conditions.

The diurnal variation obtained from the data is portrayed in figures 1 and 2, where the number of cases of auroral interaction is plotted against time of day (75° W meridian time). Figure 1 illustrates the diurnal distribution by months, and the total distribution for the 36-month period is displayed in figure 2.

§ 3. Diurnal Distribution in the Ionized Aurora

During the term 1949–51, the ionized aurora was observed most frequently in the spring followed, in order, by winter, summer and autumn. The frequency of occurrence in January was similar to that found for July. The months with the greatest reported number of cases were, with decreasing number, April, May and October. The month with the least activity was December.

As the months progressed from January to May, radio-wave reflections from the ionized aurora were detected earlier in the day and persisted for a longer period of time. During April–June, the maximum number of reports fell in the interval 1800–2000 E.S.T. (75° W meridian time), but during autumn the maximum appeared later, from 1900 to 2300 E.S.T. In January reflections were reported only from 1800 to 0100 E.S.T. However, in May, as in October, radio contact was established from 0900 through midnight to 0300 E.S.T.

In some months the number of cases rises somewhat smoothly to a maximum value and then decreases in a similar fashion. At other times (e.g. April, August

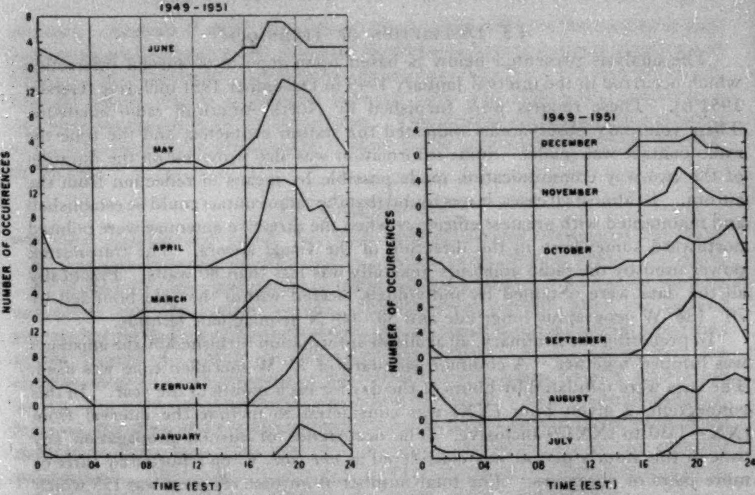


Figure 1. Diurnal variation by months, in occurrences of the ionized aurora reported over North America between geomagnetic latitude 50–60° N. Data were obtained during the three-year period 1949–1951 inclusive.

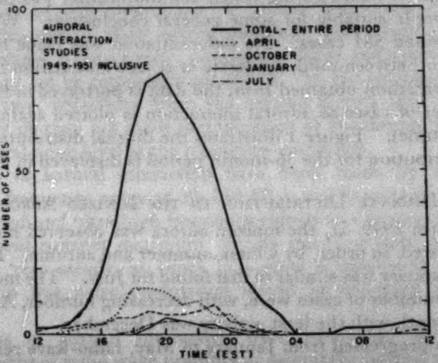


Figure 2. Diurnal variation in the occurrence of the ionized aurora considering all data obtained during the three-year interval 1949–1951. Data were obtained in the geomagnetic latitude range 50–60° N over North America.

and October) the occurrence curve evinces a bimodal distribution with the intervening minimum occurring at 2000–2100 E.S.T.

When all data were combined, a somewhat sharp maximum in the diurnal occurrence was found near 2000 E.S.T. (figure 2). Over 60% of the total occurred in the interval 1800–2200 E.S.T., and over 86% in the interval 1700–2400 E.S.T. Thus, although the ionized aurora may appear at almost any hour of the day, it seems to be primarly constrained to the diurnal period from late afternoon to early morning.

Some evidence of the existence of ionized aurorae was obtained at all hours of the day except during 0400–0600 E.S.T. However, it is believed that as a greater number of observations is accumulated, evidence of the existence of the ionized aurora during these hours will also be found. It should be noted that most of the co-operating radio amateurs did not observe during 0300–0600 E.S.T.

§ 4. COMPARISON WITH OTHER INVESTIGATIONS

An attempt was made to compare the above results on the diurnal distribution of the ionized aurora with those obtained by Currie *et al.* (1953), Hawkins (1954) and Dyce (1953). Unfortunately, the basis for the tabulation used by each investigator is somewhat different, and the various results may be intercompared only to a limited extent. Employing data which could be construed as somewhat homogeneous, the graphs of figure 3 were prepared. The diurnal variation in the occurrence of the ionized aurora, as given in each paper, was utilized.

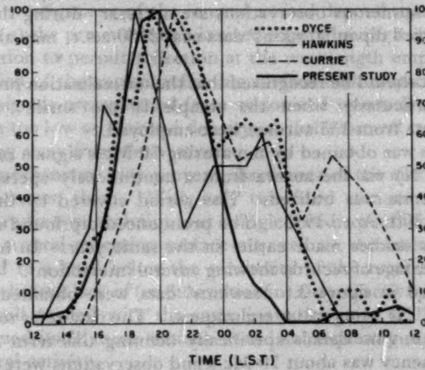


Figure 3. Comparison of diurnal variation of the ionized aurora observed within the geomagnetic latitude range 50–61° N, as reported by different investigators. Data are normalized to the maximum value for the day.

A ratio for each hour of the day was determined as the value given for the hour concerned divided by the greatest hourly value in the 24-hour period. This normalization procedure allowed the maximum for each group of observations to be plotted as 100. The information presented by Hallgren and Meos (1952), Bullough and Kaiser (1954), and Harang and Landmark (1954) could not be employed in the comparison study because of either paucity of data, or difficulty in removing the weighting factors used in their analysis.

In spite of the apparent discordance among the various normalized curves displayed in figure 3, several conclusions may be drawn:

(*a*) The maximum of activity in the ionized aurora occurs between 1800 and 2200 L.S.T.

(*b*) On the average, the existence of an ionized aurora during a given night may become known between 1200 and 1500 L.S.T., when activity first begins to increase progressively towards the nocturnal maximum.

(*c*) The maximum in the number of occurrences is attained about six hours after the appearance of, or an increase in, auroral interaction is first noted.

(*d*) In general, the number of occurrences decreases after 2200 L.S.T. Distortions and dips in individual curves may possibly be ascribed to an insufficient number of observations. These may be smoothed out when more data are obtained.

(*e*) Tentatively, the curve of diurnal variation in the ionized aurora does not seem to be too dependent upon geomagnetic latitude in the range 50° N–60° N, especially from 1200 to 2200 L.S.T.

(*f*) Data (Currie *et al.* 1953) from the highest geomagnetic latitude station, Saskatoon, seem to indicate that the ionized aurora may be observed later in the day, and may persist until later the following day than at other stations. This conclusion is very tentative and before being fully accepted must be supported by additional observations at higher geomagnetic latitude stations.

(*g*) Marked irregularities occur in the curves during the morning hours from 0100 to 1200 L.S.T. The sharp breaks in some curves may arise from a paucity of data. More numerous observations are necessary during these periods.

(*h*) The marked dip in Hawkins' data near 2100 L.S.T. may arise from too few observations.

In general, it should be recognized that the normalization process emphasizes fluctuations, particularly when the sample is too small. In the present investigation, data from 155 aurorae were employed.

Dyce's curve was obtained by monitoring 50 Mc/s signals received at Ithaca, U.S.A. ($\Lambda = 54°$ N) via the aurora from a continuously operating transmitter. A directive antenna was utilized. The period covered by the curve is from 1st February to 30th April 1953. The pronounced dip found around midnight also appeared in studies made earlier in the same year. In his analysis Dyce refers to a 'percentage of records showing auroral interaction'. It is these results which are plotted in figure 3. Hawkins' data were obtained at Manchester, England ($\Lambda = 56°$ N) using radar techniques. The data refer to a 'frequency of auroral echoes' but no details specifically defining this term are given. The radio wave frequency was about 70 Mc/s and observations were made some time between September 1949 and about December 1950. Data from 19 aurorae constitute the basis of the results. Currie's material was also obtained by radar methods at Saskatoon, Canada ($\Lambda = 61°$ N) using a frequency of 56 Mc/s. The study period included the interval 23rd February to 9th April 1952. A 'frequency of occurrence' of auroral interaction is defined for each hour; this classification is a measure of the percentage of the number of two-minute intervals with echoes to the number of two-minute intervals that the radar had been operated. In their paper, Currie *et. al.* stated that "data for the daylight hours ... are not as extensive as for the hours of darkness and twilight, lack of personnel at that time making operation impossible throughout the 24-hour period".

It would have been interesting to compare the results here with those of Hallgren and Meos. Unfortunately, however, their 'frequency of occurrence' contains a weighting factor which cannot be readily eliminated from their data to place them on a comparable basis with those shown here. It should be noted that some bias may exist in the present data. Undoubtedly most of the volunteer observers were inactive from 0300 to 0600 L.S.T., and the curve may be expected to be deficient during this interval. This expectation is partially confirmed in figure 3, where the curve based on the present study attains its minimum at 0400 E.S.T., several hours prior to the minimum found by other investigators. On the other hand, considering the large number of cases which comprise the present study, it is possible that the *relative* number of occurrences at 0300 L.S.T. is in reality quite small. To clarify the situation it would be desirable to obtain additional data continuously during the early morning period.

§ 5. DISCUSSION OF RESULTS

The diurnal variation in the occurrence of the ionized aurora seems to be similar to that of the visual aurora. The maximum in occurrence of the former is between 1800 and 2200 L.S.T., whereas for the visual aurora the maximum in intensity is between 2200 and 0200 L.S.T. (Currie and Edwards 1934). Chapman and Bartels (1940) note that in the non-polar regions of Europe and North America, aurorae are most frequent during the early night hours with a maximum at about 2200 L.S.T.

Radio-wave investigations of the ionized aurora depend upon the fulfilment of two criteria: (*a*) the presence of an ionized aurora having a sufficient electron or ion concentration to permit reflection at the wavelength employed, and (*b*) a suitable orientation of the auroral reflecting surface with respect to the incident radio ray. No signals are returned from an ionized aurora if no aurora exists, if one exists with too low an electron density, or if the auroral curtain directs the radio-wave energy away from the observer.

While both mechanisms may prevent observance of the aurora, it is possible to distinguish between them. With only two fixed stations (transmitter and receiver) or with radar type techniques at a single station, a sufficient change in the direction of the auroral reflecting surface may reduce the intensity of returned (scattered) signal to a level below the sensitivity of the receiver. Thus, these stations would indicate a lack of (or diminution in) auroral activity even though the ionized aurora still existed with the same ionic density. If then the shift in orientation reversed itself, these stations would subsequently report an apparent increase in auroral activity.

On the other hand, the grid reporting system under the same conditions may show no decrease in auroral occurrence. The network must be large enough so that a shift in orientation of the auroral refracting screen does not deflect the radio ray outside the geographic area of the grid. With a given receiver a change in direction of the reflector would shift the area of observation. New stations would be contacted by the given station as the auroral reflector changed its orientation. The grid reporting system then could report the existence of an ionized aurora irrespective of the angle of incidence of the radio ray; one or two fixed stations, however, may not observe an ionized aurora if its orientation is unfavourable.

The bimodal distribution obtained by Hawkins and Currie and mentioned by Dyce may arise from a shift in orientation of the ionized aurora, as described above. The absence of a double maximum in the amateur reports can scarcely be attributed to a lack of observers prior to 0200 L.S.T., and gives some credence to the possibility that the orientation change reduces the observability of the ionized aurora when a single station or two stations are employed. This change in orientation may coincide with the change from one auroral form to another. These latter conclusions must be accepted with reserve until more extensive studies utilizing a continuously operating network are undertaken.

The low occurrence of the ionized aurora around local noon may possibly be explained by assuming (*a*) ultra-violet or (*b*) geomagnetic control over the aurora. The former possibility would involve an activation or deactivation of the atmospheric particles by solar radiation so that a given energy stream of corpuscles no longer ionizes the atmospheric molecules as efficiently as before. In the second case, the intensity of the geomagnetic field and the concentration of the lines of force may be different over the sunlit than over the dark side of the earth. With such an asymmetrical distribution in the lines of force (under which the earth turns), the aurora may be permitted to extend to lower latitudes in the nocturnal than in the daylight hemisphere. It seems more probable, however, that the magnetic rigidity of the auroral primaries is such that they cannot penetrate to the earth's atmosphere near the subsolar point. If so, this fact should provide some indication of the energy spectrum of the auroral primaries. The problem is by no means settled, but the speculations are interesting.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance provided this study by the Radio Amateurs of North America who on a voluntary basis obtained the data on which this study is based. The generous aid and co-operation of O. P. Ferrell in supervising the data collection is also gratefully acknowledged.

REFERENCES

ASPINALL, A., and HAWKINS, G. S., 1950, *J. Brit. Astr. Ass.*, **60,** 130.
BULLOUGH, K., and KAISER, T. R., 1954, *J. Atmos. Terr. Phys.*, **5,** 189.
CHAPMAN, S., and BARTELS, J., 1940, *Geomagnetism*, Vol. 1 (Oxford: University Press), p. 472.
CURRIE, B. W., and EDWARDS, H. W., 1934, *Terr. Magn. Atmos. Elect.*, **39,** 293.
CURRIE, B. W., FORSYTH, P. A., and VAWTER, F. E., 1953, *J. Geophys. Res.*, **58,** 179.
DYCE, R., 1953, Cornell University, *Report EE* 156, p. 12.
FERRELL, O. P., 1948, *Proc. Inst. Radio Engrs, N.Y.*, **36,** 879.
FORSYTH, P. A., 1954, AF Cambridge Research Center, *Geophys. Res. Paper* No. 34.
FORSYTH, P. A., PETRIE, W., and CURRIE, B. W., 1949, *Nature, Lond.*, **164,** 533.
GERSON, N. C., 1951 a, *Nature, Lond.*, **167,** 804; 1951 b, *Canad. J. Phys.* **29,** 251.
HALLGREN, G., and MEOS, J., 1952, *Tellus*, **4,** 249.
HARANG, L., and LANDMARK, B., 1954, *J. Atmos. Terr. Phys.* **4,** 322.
HARANG, L., and STOFFREGEN, W., 1940, *Hochfrequenztech. u. Elektroakus.*, **55,** 105.
HAWKINS, G. S., 1954, Trans. Brussels Meeting, International Association for Terrestrial Magnetism and Electricity, Bull. No. 14, p. 229.
LOVELL, A. C. B., CLEGG, J. A., and ELLYETT, C. D., 1947, *Nature, Lond.*, **160,** 372.
MCKINLEY, D. W. R., and MILLMAN, P. M., 1953, *Canad. J. Phys.*, **31,** 171; 1954, AF Cambridge Research Center, *Geophys. Res. Paper* No. 34.
MOORE, R. K., 1951, *J. Geophys. Res.* **56,** 97.
PETERSON, A. M., and LEADABRAND, R. L., 1954, *J. Geophys. Res.* **59,** 306.
WILSON, M. S., 1941, *QST.*, **25,** No. 8, p.23, No. 9, p.23.

Demagnetization Experiments on Cerium Ethyl Sulphate

By A. H. COOKE, S. WHITLEY† and W. P. WOLF
Clarendon Laboratory, Oxford

MS. received 27th January 1955

Abstract. At temperatures below 1°K the magnetic susceptibility of cerium ethyl sulphate is highly anisotropic, the molar Curie constant being 1·35 parallel to the crystal axis and less than 0·01 perpendicular to the axis. Demagnetization experiments using fields along the axis show that the magnetic susceptibility in this direction passes through a maximum at low temperatures, and that the specific heat is too large to be accounted for by magnetic dipole interaction between the magnetic ions. This behaviour is attributed to antiferromagnetic coupling of the ions, possibly due to the atomic electric quadrupole–quadrupole coupling suggested by Finkelstein and Mencher.

It is shown that after demagnetization to a low temperature it is possible to apply a magnetic field in a direction of low susceptibility without greatly raising the temperature, so that magnetic measurements can be made below 1°K. This technique has been applied to the determination of the critical fields of superconducting zinc and lead at temperatures below 1°K.

§ Introduction

In cerium ethyl sulphate, the $^2F_{5/2}$ ground state of the free Ce^{3+} ion is split by the crystalline electric field into three doublets which can be characterized approximately by the component of angular momentum parallel to the hexagonal axis of the crystal, $J_z = \pm 1/2$, $J_z = \pm 3/2$, and $J_z = \pm 5/2$ (Elliott and Stevens 1951, 1952.) From measurements of the magnetic properties of the salt in the helium temperature range, it is found that the $\pm 5/2$ doublet is the lowest energy level, and that the $\pm 1/2$ doublet is separated from it by an energy Δ where $\Delta/k = 7°$ approximately (Bogle, Cooke and Whitley 1951). As the $\pm 3/2$ doublet is separated from the ground level by an energy of about 180°, the magnetic properties of the salt in the helium temperature range depend entirely on the two lower doublets, and the specific heat in this range is dominated by the Schottky anomaly due to transitions from the $\pm 5/2$ level to the $\pm 1/2$ level as the temperature is raised. From susceptibility and paramagnetic resonance measurements, it is found that the properties of these two levels can be expressed by the spectroscopic splitting factors $g_{\parallel}$ and $g_{\perp}$ respectively parallel and perpendicular to the crystalline axis. For the $\pm 5/2$ level, $g_{\parallel} = 3{\cdot}80 \pm 0{\cdot}04$ and $g_{\perp} = 0{\cdot}22 \pm 0{\cdot}04$, while for the $\pm 1/2$ level, $g_{\parallel} = 1{\cdot}0 \pm 0{\cdot}2$ and $g_{\perp} = 2{\cdot}2 \pm 0{\cdot}2$. The results of the present work give $6{\cdot}6° \pm 0{\cdot}1°$ for the splitting Δ/k of the two levels.

At temperatures below 1°K only the $\pm 5/2$ level will be appreciably populated, and the magnetic susceptibility of the salt will be highly anisotropic. Ignoring temperature-independent effects, the susceptibility is proportional to g^2 and is therefore more than a hundred times greater along the crystalline axis than perpendicular to it. In an earlier paper (Bogle, Cooke and Whitley 1951) it was suggested that cerium ethyl sulphate would be a particularly interesting substance for adiabatic demagnetization work, since after cooling the salt by demagnetization using a field parallel to the crystal axis, it would be possible to apply a field perpendicular to the axis without greatly raising the temperature. The salt would then be a valuable cooling agent for experiments requiring the application of a magnetic field below 1°K, for example the tracing of the transition curve of a superconductor, or paramagnetic or nuclear resonance measurements. In this paper we give an account of experiments on the cooling of single crystals of cerium ethyl sulphate by adiabatic demagnetization, and on the effect of magnetic fields perpendicular to the crystalline axis.

---

† Now at U.K.A.E.A. Capenhurst Works, Chester.

§ 2. DEMAGNETIZATIONS TO ZERO FIELD

Four single crystals grown from aqueous solution and weighing in all 2·44 grammes were mounted together with their axes vertical and suspended by thin nylon threads inside a double-walled cryostat of Pyrex glass. The demagnetizations were made in the usual way, the specimen being first cooled to 0·95°K by evaporation of liquid helium contained in the outer annular space of the cryostat, and isolated after magnetization by evacuating the inner space of the cryostat. The vertical magnetic field was provided by the iron-free solenoid described by Daniels (1950). The maximum field used was 7 kilo-oersteds, corresponding to an input of 25 kw to the solenoid. The magnetic susceptibility of the specimen along the crystal axis was measured by its effect on a mutual inductance coil surrounding the specimen, with its primary energized by 33 c/s a.c. producing a vertical oscillatory field of 3 oersteds at the specimen. The mutual inductance was measured by a Hartshorn bridge. In each experiment the mutual inductance was first measured with the specimen at different known temperatures down to 1°K. In the range 1°K to 2°K the susceptibility follows Curie's law, $\chi = C/T$, and extrapolation of this law to temperatures below 1°K gives the 'magnetic temperature' T^* of the salt (Kurti and Simon 1938, Cooke 1949). No correction was made for the effect of the demagnetizing and Lorentz interaction fields in the crystal as this was very small. The first experiments consisted of a series of magnetizations in known temperatures and fields, followed by measurement of the value of T^* reached on adiabatic demagnetization to zero field. The results are shown in table 1. The table shows however that the magnetic temperature

Table 1

x	0·284	0·354	0·423	0·564	0·705	0·848	0·988	1·13	1·27	1·42	1·68	2·05
$\delta S/R$	0·0100	0·0159	0·0219	0·0382	0·0582	0·0824	0·1085	0·137	0·167	0·201	0·256	0·338
T^*	0·745	0·274	0·208	0·133	0·107	0·0964	0·0905	0·0867	0·0832	0·0820	0·0806	0·0814

The parameter x given in the first row of figures is defined by $x = g_{\parallel}\beta H/kT$, where β is the Bohr magneton, H is the magnetic field and T is the temperature of magnetization and k is Boltzmann's constant. This parameter determines the decrease of entropy δS on magnetization, which is given in the second row in units of the gas constant R. The third row of figures gives the magnetic temperature T^* reached on demagnetization. The greater the entropy extracted, the lower will be the absolute temperature reached on demagnetization.

T^* passes through a minimum value, that is, the magnetic susceptibility passes through a maximum as the temperature falls. This behaviour is characteristic of substances in which the interactions between the magnetic ions are antiferromagnetic. At temperatures well removed from the susceptibility maximum the interactions will give rise to a specific heat varying as $1/T^2$. At temperatures in this range the entropy in zero magnetic field will be given by $S/\mathbf{R}=\text{constant}-A/2T^2$.

In figure 1 is given a graph of entropy against $1/T^{*2}$. It will be seen that at high temperatures the graph is linear. Assuming that at these temperatures the magnetic temperature T^* does not differ appreciably from the absolute temperature T, we obtain from the initial slope of the curve a value for the specific heat of the salt in zero field $c=A\mathbf{R}/T^2=11{\cdot}2\times10^{-4}\mathbf{R}/T^2$.

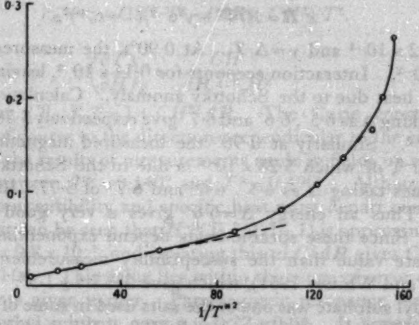


Figure 1. Demagnetizations of cerium ethyl sulphate. δS = entropy extracted by magnetization. T^* = ‘magnetic temperature’ attained on demagnetization.

This value is six times that to be expected from magnetic dipole interaction between the ions (Daniels 1953). It follows that in this salt some other interaction effect between the magnetic ions is important. This behaviour contrasts with that of neodymium ethyl sulphate, the specific heat of which can be wholly accounted for by nuclear hyperfine interaction and magnetic interaction between the ions (Roberts, Sartain and Borie 1952). (There is no nuclear hyperfine interaction in cerium salts, as none of the cerium isotopes has a nuclear spin.) Finkelstein and Mencher (1953) have suggested that the anomalous behaviour of cerium ethyl sulphate may be due to interactions between the atomic electric quadrupole moments of the magnetic ions. Since cerium is the first member of the rare earth series, the radius of the 4f orbit, which enters their specific heat formula as r^{16}, is particularly large. It is therefore very possible that this interaction makes an important contribution to the specific heat of this salt, although its effect is negligible in other cases. Because the theoretical expression for the specific heat depends so critically on this radius, and also on the energy of the $\pm 3/2$ state, which is not known at all exactly, it is not at present possible to make an accurate comparison between theory and experiment, but it is clear that the effect is of the correct order of magnitude.

It might be expected that the graph of figure 1 would pass through the axis $\delta S=0$ when $T^*=0{\cdot}95^\circ\text{K}$, the temperature of magnetization, whereas in fact when $1/T^{*2}=0$, $\delta S/\boldsymbol{R}=0{\cdot}009$. The reason is that at the temperature of magnetization the upper energy level $J_z=\pm 1/2$ is still appreciably populated, whereas after demagnetization it is drained completely. The entropy associated with this change is then the same for all the demagnetizations plotted in the graph, and displaces the entropy curve by a fixed amount. From this displacement, and the known temperature of magnetization, 0·95°K, the energy separation Δ of the upper level can be calculated, and proves to be 6·5°K above the ground level. This can be compared with the value calculable from specific heat measurements made at temperatures near 1°K by the paramagnetic relaxation method. At these temperatures the specific heat arises partly from interaction between the ions and partly from the Schottky anomaly due to the filling of the upper level as the temperature is raised, i.e.

$$c/\boldsymbol{R}=A/T^2+y^2e^{-y}/(1+e^{-y})^2,$$

where $A=11{\cdot}2\times10^{-4}$ and $y=\Delta/T$. At 0·90°K the measured specific heat was $c/\boldsymbol{R}=3{\cdot}64\times10^{-2}$. Interaction accounts for $0{\cdot}14\times10^{-2}$, leaving $c/\boldsymbol{R}=3{\cdot}50\times10^{-2}$ as the specific heat due to the Schottky anomaly. Calculations of the Schottky specific heat taking Δ as 6·5°, 6·6° and 6·7° give respectively $3{\cdot}78\times10^{-2}$, $3{\cdot}52\times10^{-2}$ and $3{\cdot}24\times10^{-2}$. Similarly at 0·98° the measured magnetic specific heat was $c/\boldsymbol{R}=5{\cdot}40\times10^{-2}$, of which $5{\cdot}28\times10^{-2}$ is due to the Schottky anomaly, against calculated values taking Δ as 6·5°, 6·6° and 6·7° of $5{\cdot}77\times10^{-2}$, $5{\cdot}38\times10^{-2}$ and $5{\cdot}00\times10^{-2}$. Thus an energy $\Delta=6{\cdot}6°$ gives a very good fit with all these experiments. Since these specific heats depend exponentially on Δ, they give a more accurate value than the susceptibility measurements reported earlier (Bogle, Cooke and Whitley 1951), which gave $\Delta=7{\cdot}5\pm1{\cdot}5°$.

Cerium ethyl sulphate was one of the salts used in some of the earliest experiments on adiabatic demagnetization (de Haas, Wiersma and Kramers 1934, de Haas and Wiersma 1935). These demagnetizations were made from temperatures at which the entropy associated with the occupation of the upper energy level is rather large. They were made using powdered specimens and so suffer from the disadvantage that differently oriented grains would cool to different temperatures—a particularly important point since the magnetizing and temperature-measuring fields were perpendicular. Even if it is assumed that the grains were randomly oriented, and that after demagnetization they rapidly came to a common temperature, it is still not possible to compute the entropy change accurately, on account of the unknown entropy of mixing as the grains come to a common temperature.

§ 3. EFFECT OF A FIELD PERPENDICULAR TO THE AXIS

To measure the effect of a field perpendicular to the magnetic axis of the crystal, a demagnetization was made after prolonged pumping of the inner chamber of the cryostat to secure a high thermal insulation of the specimen. After establishing that the specimen was warming at a suitably slow rate, the solenoid was removed, and an electromagnet was wheeled into position to provide a horizontal field. The replacement of the vertical solenoid by the magnet considerably influenced the mutual inductance of the temperature-measuring coil; in a subsidiary experiment we verified that the only effect was to change the mutual inductance by a fixed amount, independent of the temperature of the specimen. The change in the axial susceptibility of the salt under the influence of fields up to 1,800 oersteds perpendicular to the crystal axis was then observed at different temperatures as the specimen slowly warmed to the temperature of the cryostat.

If we assume that the sole effect of the field is to raise the temperature of the specimen, as is to be expected since the oscillatory measuring field and the steady field are perpendicular, we have for the adiabatic change of temperature with field

$$\left(\frac{\partial T}{\partial H}\right)_S = -\frac{T}{c_H}\left(\frac{\partial I}{\partial T}\right)_H,$$

where c_H is the specific heat at constant field, and I is the specific magnetization. If the temperature is not too low, we may assume that Curie's law $I = CH/T$ is obeyed, and that

$$c_H = c_0 + CH^2/T^2 = (AR + CH^2)/T^2.$$

Then

$$\left(\frac{\partial T}{\partial H}\right)_S = \frac{CH}{AR + CH^2}$$

and on integrating $(T_H/T_0)^2 = 1 + CH^2/AR$. The Curie constant C in this expression refers of course to the direction perpendicular to the crystal axis.

Table 2 shows the results of measurements made in fields up to 1800 oersteds at temperatures between $T^* = 0{\cdot}140°$ and $T^* = 0{\cdot}490°$. Although at the lower temperatures the susceptibility and specific heat must depart considerably from the laws assumed it can be seen that $H^{-2}(T_H{}^2/T_0{}^2 - 1)$ is approximately constant. Taking its value as $9{\cdot}8 \times 10^{-7}$, and equating this to C/AR gives $C = 0{\cdot}0090$, from which $g_\perp = 0{\cdot}31 \pm 0{\cdot}04$. This value lies within the range of error of the susceptibility measurements given earlier. Paramagnetic resonance measurements by Bogle on cerium ethyl sulphate gave $g_\perp = 0{\cdot}22 \pm 0{\cdot}04$. It appears then that the effect can be described simply as a heating. In small fields the heating is given by $\delta T/T = 5 \times 10^{-8} H^2$; for example a demagnetization from a temperature of 0·95°K and a field of 1800 oersteds along the crystal axis cools the salt to 0·20°. If the same field is applied perpendicular to the axis the salt is heated only to 0·23°. In effect it has been cooled by rotation of the field.

Table 2

T^*	0·140	0·153	0·161	0·173	0·200	0·211	0·229	0·256	0·270	0·295	0·322	0·342	0·490
H	1·84	1·34	0·455	1·39	1·84	1·82	1·82	1·37	1·37	1·82	1·82	1·82	1·82
r	1·135	1·085	1·01	1·085	1·142	1·160	1·145	1·085	1·165	1·165	1·170	1·150	1·110
K	8·5	9·3	9·2	9·2	9·0	10·4	9·3	9·3	10·2	10·6	11·1	9·7	7·0

The first row of figures gives the temperature of the specimen, as measured by the magnetic thermometer, in zero magnetic field. The second row gives the field H in kilo-oersted which causes the magnetic temperature to rise by the ratio r shown in the third row. The fourth row K gives the value of $(r^2-1)/H^2$, in units of 10^{-8} oersted^{-2}.

§ 4. THE TRANSITION CURVES OF SUPERCONDUCTORS

As an example of the type of measurement which can be made using this salt as a cooling agent, we made a very simple measurement of the superconducting transition curve of zinc. Six crystals of cerium ethyl sulphate were glued with their axes vertical to two strips of zinc foil each 3 cm × 1 cm × 0·004 cm which were then mounted together in a vertical plane and suspended in the cryostat exactly as in the previous experiments. The foil, which was supplied by I.C.I., was of 99·95% purity. It was not annealed. After demagnetization using a vertical field had cooled the crystals and zinc to about 0·1°K the solenoid was removed and a horizontal field parallel to the plane of the zinc foil was applied by means of a pair of Helmholtz coils. The mutual inductance used to measure the susceptibility of the crystals also served to detect the destruction of superconductivity in the zinc by the apparent change in the susceptibility of the specimen, amounting to $1/4\pi$ e.m.u. per unit volume of zinc, which occurred when the zinc became normally conducting. The corresponding temperature was given by the susceptibility of the cerium ethyl sulphate, measured with the zinc in the normal state. As the critical field for zinc is only some 50 oersteds the effect of this field in raising the temperature of the specimen was quite negligible. The results of the measurements are shown in figure 2. The critical field at the lowest temperatures $H_0 = 50 \pm 3$ oersteds, and the critical temperature $T_0 = 0{\cdot}89 \pm 0{\cdot}01$°K, in good agreement with the values obtained by Goodman and Mendoza (1951). The chief inaccuracies in our measurement were caused by the rather rapid rise of temperature of the specimen above 0·5°K. This could be reduced by using a larger specimen than the 3 grammes of crystal used in this experiment.

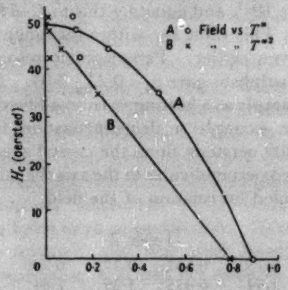


Figure 2. Critical fields for superconductivity of zinc plotted as a function of temperature (curve A) and temperature² (curve B).

To test this technique in higher magnetic fields a similar experiment was made on a specimen of lead foil, also supplied by I.C.I. The superconducting transition was first observed at different temperatures in the helium range by noting the apparent susceptibility of the specimen, as measured by the Hartshorn bridge, with the steady magnetic field held at different values in the neighbourhood of the transition value. The transition was observed at 545 oersteds at 4·25°K, at 720 oersteds at 2·18°K, and at 780 oersteds at 1·00°K. After demagnetization the transiton was observed in the same way at a temperature of 0·12°K, and was found to occur at a field of 800 ± 20 oersteds, in agreement with the value obtained by Daunt, Horseman and Mendelssohn (1939) by extrapolation from measurements made at temperatures down to 1·5°K.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr. N. Kurti, Dr. K. Mendelssohn and Dr. J. L. Olsen for their helpful advice, and to Dr. H. Meyer for his assistance in the experiments. This research was carried out during the tenure of Department of Scientific and Industrial Research awards by two of the authors (S. W. and W. P. W.).

REFERENCES

BOGLE, G. S., COOKE, A. H., and WHITLEY, S., 1951, *Proc. Phys. Soc.* A, **64**, 931.
COOKE, A. H., 1949, *Proc. Phys. Soc.* A, **62**, 269.
DANIELS, J. M., 1950, *Proc. Phys. Soc.* B, **63**, 1028; 1953, *Proc. Phys. Soc.* A, **66**, 673.
DAUNT, J. G., HORSEMAN, A. H., and MENDELSSOHN, K. A. G., 1939, *Phil. Mag.*, **27**, 754.
ELLIOTT, R. J., and STEVENS, K. W. H., 1951, *Proc. Phys. Soc.* A, **64**, 932; 1952, *Proc. Roy. Soc.* A, **215**, 437.
FINKELSTEIN, R., and MENCHER, A., 1953, *J. Chem. Phys.*, **21**, 472.
GOODMAN, B. B., and MENDOZA, E., 1951, *Phil. Mag.*, **42**, 594.
DE HAAS, W. J., and WIERSMA, E. C., 1935, *Physica*, **2**, 335.
DE HAAS, W. J., WIERSMA, E. C., and KRAMERS, H. A., 1934, *Physica*, **1**, 1.
KURTI, N., and SIMON, F. E., 1938, *Phil. Mag.*, **26**, 849.
ROBERTS, L. D., SARTAIN, C. C., and BORIE, B., 1952, *Rev. Mod. Phys.*, **25**, 170.

The Measurement of the Dielectric Constant of Standard Liquids

By L. HARTSHORN, J. V. L. PARRY and L. ESSEN
National Physical Laboratory, Teddington, Middlesex

MS. received 8th September 1954, and in revised form 16th February 1955

Abstract. New determinations have been made of the dielectric constants of benzene, cyclohexane, dichlorethane and nitrobenzene with the object of providing values of a higher accuracy than has hitherto been available for four standard liquids covering the range of values commonly required. Liquids of specially high purity were obtained with the co-operation of Professor J. Timmermans of Brussels and Dr. Herington of the Chemical Research Laboratory, Teddington, and the measurements were made by the a.c. bridge technique developed at the National Physical Laboratory for precision capacitance measurement and also by a cavity resonator method.

Part I of the paper discusses the general problem and the special features of the capacitor designed to receive the liquids and surveys the results obtained. Part II gives details of the measurements made by capacitance methods at frequencies from 50 c/s to 10 kc/s, and Part III describes the measurements made at microwave frequencies.

The values obtained confirm estimates of the most reliable values and their probable accuracy published in 1951 by A. A. Maryott and E. R. Smith as a result of a survey of existing data made at the National Bureau of Standards, Washington. It is considered that the new values narrow the range of uncertainty by a factor varying from 3 to 10 for the different liquids.

The accuracy obtained for benzene at audio frequencies was within about 0·01%; at 9000 Mc/s the accuracy was a trifle worse but the results show clearly that although the dielectric constant exceeds the square of the optical refractive index by about 1%, no measurable diminution in dielectric constant occurs in passing from 50 c/s to 9000 Mc/s. A considerable increase in power absorption is however noticeable at 9000 Mc/s and an even greater increase at 25 000 Mc/s, and it is shown that this absorption is not entirely dependent on the water that most samples of benzene contain and that can only be removed by weeks of drying treatment. The necessity of applying such treatments is a serious disadvantage to the use of benzene as a standard liquid for measurements of dielectric constant, and cyclohexane, which is free from this disadvantage, is recommended instead.

PART I. THE GENERAL PROBLEM

By L. Hartshorn

§ 1. Introduction

This investigation arose from a suggestion of Professor J. Timmermans that there was a need for the determination of the dielectric constants of at least two more standard liquids with an accuracy comparable with that obtained by Hartshorn and Oliver (1929) for pure benzene. Liquids of low dielectric constant can be very satisfactorily measured by comparison with benzene, but for work on liquids of high dielectric constant standard liquids having values of say 10 and 40 as against 2 for benzene are badly needed. When various pure liquids are employed as standards for such work their published values frequently do not appear to be consistent with one another to the accuracy now required and it is almost impossible to decide which is the most reliable one. The uncertainties may arise from imperfections in the measuring technique or in the purity of the standard liquids, and Professor Timmermans undertook to provide suitable liquids highly purified by the processes which have been specially studied in his laboratory at Brussels if they could be measured by the technique of capacitance measurement developed at the National Physical Laboratory. The two liquids selected by Professor Timmermans were dichlorethane having a dielectric constant of 10·6 at 20°c, and nitrobenzene having a value of 35·7. He also pointed out that although no evidence had appeared since 1929 to suggest that the value then obtained for benzene was not accurate within the limits claimed by Hartshorn and Oliver (± 2 parts in 10 000), yet their uncertainty was mainly associated with the difficulty of chemical purification. With modern purification techniques there was a good prospect of producing better samples of pure benzene than those available in 1929, and therefore there would be considerable point in a re-determination of the value for benzene, if it could be included in the investigation. The experiments were therefore planned to include a new determination of the value for benzene, on the understanding that Professor Timmermans would provide samples of the required purity. The work was long delayed by the War, and by the time the measurements began Dr. Coulson and Dr. Herington at the Chemical Research Laboratory had developed techniques for producing hydrocarbons, including benzene, of extremely high purity. Arrangements were therefore made to include a sample of their benzene as well as Professor Timmermans'; Dr. Herington also gave valuable advice on the manipulation of the samples throughout the work. There can be little doubt therefore that the chemical procedure in the present work was in accordance with the best modern practice and greatly superior to that followed in 1929.

The technique of capacitance measurement has also improved considerably since 1929. The measurement of dielectric constant consists of the measurement of the ratio of two capacitances; in 1929 it was considered that in order to determine this ratio to 1 part in 10000, while excluding any capacitance associated with connecting leads or solid insulators, it was necessary that the smaller capacitance (with air or vacuum as dielectric) should have a value of the order of 1000 pF. Subsequent experience on the measurement of the capacitances of screened-grid valves and the like has shown that with adequate screening the stray capacitances of leads, etc. can be excluded from a measurement with an uncertainty less than 0·01 pF. It has therfore become possible to obtain the same accuracy of capacitance measurement with a much smaller capacitor. This is extremely important in the present work, for it permits the use of a smaller sample of the liquid and it also makes possible the use of a noble metal for the electrodes of the capacitor without making the cost unreasonably high. Both these factors, a small sample and platinum electrodes, are a great asset when chemical impurities are the chief cause of uncertainty, as in this investigation.

Proceeding from these considerations, the first step was the construction of a capacitor or cell having a precisely determined capacitance of not less than 50 pF when empty, and the means of filling it with a given sample of a pure liquid and subsequently evacuating it without affecting either the value of the capacitance when empty or the purity of the liquid.

§ 2. The Capacitor Cell

The construction of a cell including all the essential features proved to be unexpectedly troublesome. Several designs incorporating electrodes in the form of metal films deposited on the opposing surfaces of a double-walled cylindrical glass vessel were tried but had to be abandoned because of imperfections either in the films themselves, or more usually in the contact between the film and its connecting lead, which must be sealed into the glass wall. It is essential that the lead and electrode shall constitute a sensibly perfect electrical conductor, of a precisely defined geometrical form, but after several failures to satisfy this condition with films of silver deposited chemically, and of platinum and silver made with various commercial metallizing preparations, it was concluded that it would be necessary to use electrodes of solid platinum foil. The cell shown diagrammatically in figure 1 was then constructed and ultimately proved to be very satisfactory. It will be convenient to describe in turn the principal mechanical. thermal, chemical and electrical features.

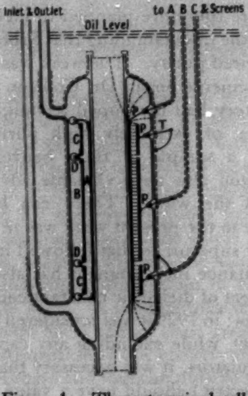


Figure 1. Three-terminal cell.

2.1. *Mechanical*

The electrodes are seamless cylinders of platinum foil 0·3 mm thick, this being considered to be the smallest thickness that would provide the required rigidity in self-supporting cylinders of the required size. The arrangement consists essentially of two coaxial cylinders, the inner one having a diameter of approximately 25 mm and the spacing between their walls being about 1·3 mm (figure 1). The outer cylinder is divided by circumferential gaps, which were made as narrow as was practicable, into three parts, a central portion B of length 10 cm forming the electrode proper, and the two end portions C, each of length about 2 cm, which serve as guard rings. The electrode B is supported in its correct position with reference to the guard rings by six small glass beads, two of which are shown in figure 1. The beads are each fused to the ends of two small lugs of platinum foil 1·5 mm wide that are welded to the cylinders at points uniformly spaced around their rims. The complete outer cylinder CBC is supported coaxially with A by two similar sets of three glass beads fused to neighbouring platinum lugs at the ends of the complete system. Two of these beads are also seen in figure 1. The beads were made of soda glass since it gives the best seal with platinum and its insulating properties are quite adequate.

The electrode assembly is mounted as shown in figure 1 within a double-walled cylindrical cell made of Pyrex glass. The whole assembly is carried by the glass tube forming the inner wall of the cell, the inner electrode A being made to fit this tube tightly. A lead connected to each of the four electrodes was brought through the outer wall of the cell. The portion of each lead within the cell was of platinum wire P, which was welded to the platinum cylinder. The portion passing through the cell wall was a short length of tungsten wire T welded to P at one end and to a length of screened cable at the other. The Pyrex glass makes a good seal with the tungsten and this seal was extended so as to cover the junction of the platinum and tungsten. Thus any liquid introduced into the cell makes contact with no metal except platinum.

The electrode structure is very flimsy in comparison with those used in the earlier work, but it has the advantage that the electrodes are comparatively free from mechanical stress and in addition the cylindrical form ensures that any slight mechanical displacement of the two working electrodes A and B relative to one another causes a minimum change of capacitance. Constancy of capacitance during the course of a sequence of measurements in which the cell is measured, first empty, then filled and then empty again, is the essential requirement and it may be said that the values obtained for the empty cell during any one day were always found to be constant to better than 1 part in 10^4. The value was 60·2 pF. During a period of a few months in which the cell was subjected to many cleaning operations the extreme deviations from the mean value amounted to ± 3 parts in 10^4. These measurements were made with the cell evacuated and it is an important feature of the design that the slight displacement of the cell walls that necessarily occurs on evacuation does not affect the geometrical capacitance of the system.

The volume of liquid required for a measurement was about 125 ml., but it should be observed that the measured capacitance is independent of the volume of liquid provided that the electrode B is completely immersed.

2.2. *Thermal*

The inlet and outlet tubes and the leads can be extended vertically upwards from the cell as far as may be necessary and therefore the cell can be completely immersed in an oil bath, the temperature of which is controlled to the required accuracy. The oil level is indicated in figure 1, from which it will be seen that the oil circulates around both the inner and outer walls of the cell. Temperature gradients within the system are therefore easily made negligible and a thermometer immersed in the liquid gives the temperature of the cell with the accuracy required.

It was considered desirable to make measurements at temperatures from 20° to 30°C, and it was therefore important that the temperature coefficient of capacitance of the cell should be small and definite, that is to say, the variation of capacitance with temperature should be free from the hysteresis observed in most commercial capacitors. If all the material composing the electrodes and separators has a coefficient of linear expansion α, and if changes of temperature impose no mechanical stresses on the system, then the coefficient of capacitance will also be α. The electrode assembly itself is composed of nothing but platinum and soda glass, both of which have a coefficient of linear expansion of about 9×10^{-6}. It was therefore expected that the cell would at best have a temperature coefficient of capacitance of this value. The value actually obtained from measurements of capacitance of the evacuated cell over the range 20° to 30°C proved to be zero, certainly less than 2×10^{-6}. The explanation is to be found in the fact that the inner electrode A fits fairly tightly the supporting Pyrex tube. Thus the diameter of electrode A is always controlled to some extent by the diameter of the Pyrex tube. If the fit were perfect the radial thermal expansion of the electrode A would be that of Pyrex, 3×10^{-6}, instead of that of platinum, but since the fit can only be such as to give contact at a few points, the effective coefficient of expansion of the inner electrode will be merely a little less than the 9×10^{-6} of freely supported platinum. If r is the radius of this electrode, s the electrode separation, α_0 the coefficient of axial expansion of both electrodes, α_1 and α_0 the coefficients of circumferential expansion of the inner and outer electrodes respectively, it is easy to see that the coefficient of capacitance will be approximately $\alpha_0 - r(\alpha_0 - \alpha_1)/s$. In the present cell r/s is approximately 10 and thus the temperature coefficient of capacitance becomes zero when $\alpha_1 = 0{\cdot}9\,\alpha_0$. It is clear therefore that perfect temperature compensation was only obtained by sheer good luck, though some compensation may reasonably be expected in a cell of this design.

2.3. *Chemical*

As mentioned earlier, it was known that a limiting factor would be the difficulties of ensuring that the dielectric filling the cell is chemically pure at the time the capacitance is measured. In 1929 it was found that the purest benzene obtainable invariably contained sufficient water to affect its dielectric constant appreciably and that this water could only be removed by months of treatment with drying agents. In the present work, although the highly purified benzene was supplied in sealed ampoules, the measurements again showed that the water content was appreciable and that long treatment with drying agents in the sealed enclosure containing the cell was an essential preliminary to a satisfactory measurement. Work on standard air capacitors and electrostatic instruments generally carried out since 1929 had also shown that air capacitors with aluminium or brass plates show imperfections in their electrical properties (appreciable power factor and variation of capacitance with frequency) that reveal the existence of non-metallic films coating the surfaces of the structure, presumably oxides, absorbed water vapour, etc. Such effects can only be eliminated with certainty by the use of a noble metal for the electrodes and by thorough evacuation of the system after chemical cleaning. The present cell with its platinum electrodes and the possibility of evacuation, sealing off, and subsequently filling and emptying the cell by low-temperature distillation within the sealed enclosure, meets all the requirements. The inlet and outlet tubes facilitate the preliminary chemical cleaning when the cell is mounted in its fixed position in the temperature-controlled bath; each of these tubes is sealed to a reservoir, one serving to contain the liquid during the drying treatment and evacuation of the cell and the other serving as receiver of the liquid during the distillation. An important point is that the cell is not subjected to the slightest mechanical disturbance throughout the sequence of measurements constituting a determination of dielectric constant.

2.4. *Electrical*

The essential electrical requirement is that the capacitance measured shall be located wholly within a vacuous dielectric medium, which is subsequently completely filled by the liquid. Thus it is essential that every element of capacitance located in the glass beads and walls shall be excluded from the measured value, as also shall every element of capacitance (between either leads or electrodes) that is external to the liquid when the cell is filled. The method employed at the National Physical Laboratory for measurements of capacitance in the range 10 to 10 000 pF is that of the Schering bridge with a Wagner earth connection and the cell was designed for use with this equipment. If the electrodes A and B are connected to the bridge arm with B on the earth-potential side of it, and C is connected to the earth terminal on the Wagner arm, then at balance B and C are at the same potential and the electric field of the system is as shown diagrammatically on the right-hand side of figure 1, the familiar guard-ring configuration. The whole of the outer walls of the cell were coated with metal foil (indicated in figure 1 by a broken line) which was connected to the screens of the two cables and to the guard rings.

With this system of connections the capacitance measured by the bridge is the direct capacitance between A and B, which is that associated with the lines of force shown as thin continuous lines in figure 1. It will be observed that the platinum supporting lugs are bent outwards so that the glass beads are well clear of this field. Indeed both the glass beads D and the outer cylindrical glass wall are in a region bounded by BC and the outer screen, all of which are at the same potential. All this glass is therefore in a region of sensibly zero field, where it cannot influence the measurement. There will however be a minute stray field from A through the narrow gaps between B and the guard rings. This stray field is however mainly between A and the screen and is therefore eliminated from the measured value along with the capacitance between AC, and the capacitance between the ends of A and the outer screen, represented in figure 1 by broken lines of force. All the capacitances associated with the leads external to the cell are to screen, and are therefore eliminated in the same way. The capacitances between the internal platinum leads P connected to A and B respectively will be included in the measured value, but this is properly located first in a vacuum and then in the liquid so that it introduces no error into the measured dielectric constant. The only parts of the measured capacitance which could conceivably cause error are (i) that between the portions of the leads to A and B that are actually embedded in the glass seals, and (ii) that represented by lines of force that pass from A through the narrow gaps to the tips of theplatinum lugs attached to B that are embedded in the glass beads. From the geometry of the system it is obvious that both these capacitances are very small relative to the measured capacitance, and moreover that of these small capacitances only a very small portion is located in glass instead of in liquid and vacuum in succession. There is therefore good reason to believe that there will be no appreciable error due to the glass that provides the necessary solid insulation of the cell.

The bridge arm itself is provided with two shielded connectors, and balance is first obtained with A disconnected and the bridge connectors screened so that their capacitances are eliminated and the reading is dependent only on the internal capacitance distribution of the bridge, which remains unaltered throughout the measurements. The cell is then connected to the bridge in the manner already described and a second reading taken. The difference of the two readings gives the required capacitance of the cell. Two such pairs of readings, one with the cell full and one with it empty, are sufficient for a determination of dielectric constant. The procedure therefore is much simpler than that followed in 1929, though the additional balance of the Wagner arms may seem a complication. This has however long been standard practice at the National Physical Laboratory and the fact that it enables accurate readings of conductivity as well as dielectric constant to be obtained makes its use very desirable in work on dielectrics. It seemed at the outset that conductivity would be much more sensitive to impurities than dielectric constant and that therefore the measured conductivity would provide a valuable index of the purity of the liquid at the time of the measurement.

§ 3. SURVEY OF RESULTS

Details of all the measurements made are recorded in Parts II and III of the paper, but it will be convenient to survey the results here. The final values are summarized in table 1.

Table 1. Dielectric Constants at 20°c (ϵ_{20}) and Temperature Coefficients *a* for the range 20°–30°c

$$\epsilon_t = \epsilon_{20} - c(t-20)$$

	ϵ_{20}	Estimated accuracy (%)	*a*
Benzene	2·2836	0·01	0·00190
Cyclohexane	2·0250	0·01	0·00155
Dichlorethane	10·663	0·03	0·0553
Nitrobenzene	35·72	0·06	0·185

The behaviour of the cell described above was in every way equal to expectations and the values of dielectric constant of benzene and cyclohexane are considered to be correct to 1 part in 10 000, the uncertainty being mainly determined by unavoidable instability of the cell and the standard capacitors employed.

This degree of accuracy for benzene is however only obtained by using samples that have been subjected to a drying agent (powdered calcium hydride) for a period of several weeks. Samples of the highest purity obtainable without this special treatment have values that are greater by 10 parts or more in 10 000. Hartshorn and Oliver's conclusion in 1929, that their accuracy was almost entirely dependent on the difficulties of drying, is confirmed, and values obtained without long drying are seen to be useless for the purpose of arriving at the precise value for the pure material. The value obtained by Hartshorn and Oliver in 1929 for the benzene purified by them was $\epsilon_{20} = 2{\cdot}2830$, but since other samples gave values a little lower and they could not obtain completely stationary values even on prolonging the drying treatment for several months, they gave this value as an upper limit for pure dry benzene. The value now given is nearly 3 parts in 10 000 higher. The discrepancy barely exceeds the estimated experimental errors of the two determinations, and there can be no doubt that the present value is the better. It now seems probable that by the standard of the present investigation none of the samples measured in 1929 was really pure, though the sign of the discrepancy suggests that the vacuum treatment of the empty cell, the use of platinum electrodes, and the direct measurement of the capacitance with a vacuum as dielectric, instead of its deduction from a measurement with dry air as dielectric, may be the determining factors. Whatever the explanation it is very satisfactory that the samples purified by Professor Timmermans and by Dr. Herington, after complete drying, gave values which were in agreement and quite stable within 1 part in 10 000. Equally consistent values were also obtained for a sample of pure cyclohexane supplied by Professor Timmermans.

The accuracy obtained with the hydrocarbons proved to be unobtainable with the polar liquids; both dichlorethane and nitrobenzene showed considerable ionic conductivity and an instability which limited the accuracy, the error being about ± 3 parts in 10 000 for dichlorethane and ± 6 parts in 10 000 for nitrobenzene. The high dielectric constant of these liquids is an indication of a condition favourable to the dissociation of the dipolar molecules into separate ions, but Professor Timmermans reported that the liquids showed negligible conductivity when first prepared and it therefore seems likely that the ionization is caused by external means such as radiation in the locality. The effects of the ions were minimized by removing them from the liquid by the application of a large constant voltage which causes them to drift to the electrodes. This voltage was maintained continuously, a small a.c. voltage being superposed on it for the purpose of capacitance measurement when necessary. In these conditions the conductance steadily diminished and the capacitance reached a stationary value independent of the conductance, provided the latter was well below the excessive values reached on long standing of the liquid with no applied voltage. Details of these measurements are given in Part II.

A detailed comparison of the present results with those of other workers is unnecessary because the previous results have all been reviewed by Maryott and Smith (1951). Table 2 lists the values which on the available evidence they considered to be the most reliable for the standard liquids measured by us.

Table 2. Estimated Best Values in 1951 (Maryott and Smith)

	ϵ_{20}	Probable Accuracy (%)	a
Benzene	2·284	0·1	0·0020
Cyclohexane	2·023	0·1	0·0016
Dichlorethane	10·65	0·1	0·057
Nitrobenzene	35·74	0·2	0·181

It is very satisfactory to note that our values in every case confirm their estimate of the most reliable value and its probable accuracy, and that our investigation has considerably reduced the uncertainty for these four materials covering almost the entire range for which standard liquids are required.

The values given are all static values, that is to say, they hold good for all frequencies below the range in which dispersion first becomes detectable as the frequency is increased. Experiments described in Part III of the paper show that for benzene dispersion has not become measurable even at 9000 Mc/s. Apparent variations of dielectric constant with frequency for such liquids should therefore be regarded as an indication of errors of measurement.

§ 4. THE CHOICE OF METHOD

It may be worth while to consider here the relative merits of the various techniques now available for the measurement of dielectric constant and in what direction a further improvement in accuracy is most likely to be made.

The basic technique of the present investigation is the measurement of capacitance by an a.c. bridge method at low frequencies. The accuracy is necessarily limited by the stability of laboratory standards of capacitance and this limit has been reached with liquids of negligible conductivity, which appear to be capable of at least the same order of stability as the capacitance standard.

When the accuracy mentioned above is required it is best to work at a low frequency, since at higher frequencies the capacitance of a standard capacitor is subject to an appreciable correction proportional to the square of the frequency and to the inductance of the leads, both internal and external, connecting the electrodes of the capacitor to other elements of the bridge network. Such inductances can be measured approximately and it is easy to ensure that their effects are completely negligible at the frequencies employed in the measurements of Part II. However, when dealing with conducting liquids, uncertainties associated with the conductivity are liable to be larger than those arising from inductance, and an improvement in accuracy is therefore sometimes to be gained by working at a higher frequency, since the capacitance currents of the network, broadly speaking, increase in proportion to the rise of frequency, while the conductance current remains constant. When working with liquids like dichlorethane and nitrobenzene it would in general be preferable to work at frequencies higher than those used in Part II. The measurements in Part II show the same uncertainty at 10 kc/s as at 1 kc/s and we therefore conclude that the limit is set by instability of the liquid itself and that no improvement is to be expected with these liquids at still higher frequencies. Nevertheless this limiting accuracy would probably be obtained more easily at the higher frequencies; for example, the complication of long application of a constant voltage would probably not be necessary, and we therefore recommend the use of a higher frequency for general measurements on such liquids. The widespread use of resonance methods and beat methods at radio frequencies is no doubt justified in this way, but it is to be remembered that in such methods the errors associated with stray capacitance external to the various capacitors and with residual inductance, both self and mutual, in leads and capacitors may become relatively large, and that they cannot be eliminated or evaluated with a precision comparable with that obtained with bridge methods. For measurements of the highest precision bridge methods at a frequency low enough to ensure that errors due to inductance are negligible are not likely to be surpassed, so long as the basic method of determining the ratio of two capacitances is followed.

We may note in passing that the mere question of sensitivity of the measurement is no longer of great importance; so long as a frequency within the range for which amplifiers are easily constructed is adopted, a sensitivity that is more than adequate can be obtained.

The recent development of the cavity resonator as an instrument of the highest precision has however raised the question whether an accuracy of a higher order than is obtainable with a capacitor could not be obtained by measuring the resonant frequencies of a cavity when evacuated and when filled with the liquid, instead of making two measurements on a capacitor. A cavity constructed entirely of metal might well have a stability of a higher order than that of a capacitor, which must include some insulating component. For gases Dr. Essen has shown that the cavity technique does lead to greater accuracy. For liquids the question remains unsettled. The experiments described in Part III of the paper may be regarded as an attempt to obtain higher precision with liquids by the use of the cavity technique that had been found so successful for gases. The results were rather disappointing. The precision obtained was a little lower than that for capacitance methods, so that although the experiments established the very important fact that the dielectric constant of benzene at cavity-resonator frequencies is not measurably different from that at lower frequencies to the highest accuracy so far obtained by either method, they have not yet shown any promise of an accuracy for liquids higher than the limit at present set by the stability of standard capacitors.

PART II. A NEW DETERMINATION OF THE DIELECTRIC CONSTANTS OF BENZENE, CYCLOHEXANE, DICHLORETHANE AND NITROBENZENE

By J. V. L. Parry

§ 1. Filling and Drying Arrangements

The cell described in Part I having been constructed, it was decided to make first a new determination of the value for benzene and then to proceed to the liquids of higher value, possibly using the benzene as an intermediate standard. Since it was known that impurities, and particularly water, were the chief source of difficulty the first step was to clean the cell throughly and to mount it so that it could be successively filled and emptied in conditions excluding both air and moisture, and so that its capacitance could be accurately measured at any stage.

The preliminary cleaning of the cell was done with chromic acid followed by repeated washing with distilled water, pure alcohol, and finally pure benzene. The cell was then mounted in a vertical position and sealed to an all-glass vacuum distillation apparatus. Before any measurements were made the whole of this apparatus, including the cell, was continuously evacuated at a pressure of approximately 10^{-5} mm Hg for a period of one week. This treatment was considered necessary in order to remove as completely as possible all traces of the water taken up by the glass, and liable to contaminate a liquid introduced into any part of the system, in particular the distilling and receiving vessels and the cell.

It was necessary to apply a vacuum grease to the ground joints of the system in order to make it vacuum tight, but care was taken to ensure that the grease was located well within the edges of the joints and that the joints were placed in positions such that the liquids under investigation made no contact with them except in the state of vapour. In these circumstances there was no risk of contamination by the grease.

The cell was immersed in an oil-bath, thermostatically controlled so that the temperature was uniform throughout to better than ± 0·01°c. The temperature of the liquid within the cell was measured to this accuracy by means of a mercury thermometer immersed in the bath alongside the cell. Measurements of capacitance could be conveniently made at any temperature within the range 20° to 30°c, and the temperature coefficient of capacitance determined with an accuracy limited only by the accuracy of capacitance measurement. Fortunately, the cell itself when empty showed no detectable variation of capacitance over the range 20° to 30°c.

A sample of benzene to be measured was introduced into a reservoir (detachable by means of a ground conical joint) made to hold rather more than the 125 ml. of liquid required to fill the cell. The benzene was then frozen and the apparatus evacuated and sealed. Then by cooling with methylated spirit and solid CO_2 a second reservoir mounted immediately over the cell, the benzene was distilled at room temperature into this second reservoir, in which it was finally melted and run into the cell. The capacitance was then measured when the desired temperature condition had been reached. When the measurements were completed the first reservoir was cooled until the benzene was distilled back into it. There it was frozen and the apparatus once more evacuated and another measurement of the capacitance of the empty cell made. A measurement of the capacitance of the empty cell was made before and after each series of measurements on the liquid, and it is important to observe that the cell suffered no mechanical disturbance throughout the whole series of operations.

It was necessary to make provision for drying the liquid after its introduction into the evacuated system. The drying was done by powdered calcium hydride added to the liquid in the distilling vessel. The benzene was left standing over the calcium hydride for a week or so; it was then frozen and any liberated hydrogen removed from the system by evacuation, after which a series of capacitance measurements was made. This process was repeated until further drying caused no further diminution in dielectric constant.

§ 2. THE MEASURING CIRCUIT

The measurements of capacitance were made by means of the Schering bridge constructed at the National Physical Laboratory for general investigations into dielectrics. The essential features have been described in various publications and it is sufficient to state here that the arrangement used for the present work is shown in figure 2. The inner electrode A of the cell was connected to the terminal T_1, the outer electrode B to T_2 and the guard rings and all the screens of the cell and its leads were connected to earth and therefore to the point E on the Wagner arms. Thus all the capacitances associated with the screens were thrown into the Wagner arms and the direct capacitance alone of the cell was connected between T_1 and T_2. By means of the switch S_1 this capacitance could be connected across either arm 1 or arm 2 of the bridge.

The actual readings of capacitance were all made on standard air capacitors shown in figure 2 in arm 1 of the bridge. They included three continuously variable standards having ranges of 30–150 pF, 40–750 pF, and 200–3000 pF respectively, each of the two larger ones being readable with an accuracy to about 1 part in 20 000 of its full-scale value, and a fixed three-terminal standard of 1000 pF. Any parallel combination of these four capacitors could be inserted into arm 1 of the bridge by means of the switches shown. The capacitors were each calibrated by reference to another standard. It is unnecessary to enter into details, but it should be mentioned that all the measurements are made as differences of capacitance between two settings of each of the capacitors used, one obtained with the cell connected to the circuit and the other with the direct capacitance of the cell disconnected. The total capacitance at any setting is only known to a much lower order of accuracy than that obtained by calibration of such differences. The fixed standard capacitor was also calibrated as the difference of capacitance between two settings of a swivelling link, one putting the capacitor in circuit and the other disconnecting it. Since all the capacitors were individually screened, these differences of capacitance are strictly additive for two or more of the capacitors connected in parallel. The full accuracy of the equipment is therefore obtained for such combined capacitances provided that the difference measured occupies substantially the whole of the range of the largest capacitor in circuit for the two readings.

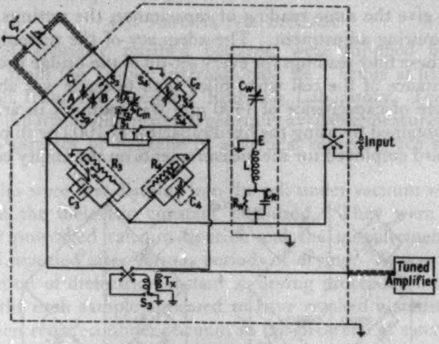


Figure 2. Schering bridge.

C_1'	Three-terminal standard air capacitor 1000 pF.
C_1 A	Two-terminal standard variable air capacitor 750 pF (or 150 pF).
C_1 B	Two-terminal standard variable air capacitor 3000 pF.
R_3 and R_4	Non-inductive ratio arms.
C_2	Variable and decades 100 pF to 0·01 μF.
C_3	Decade 100 pF to 1000 pF.
C_4	Variables 40 to 100 pF and 100 to 1000 pF.
C_m	Variable 2 to 6 pF.
C_w	Variable 100 to 1000 pF. Decade 0·0001 to 0·001 μF.
L	Variable 0·5 to 1·6 mH. Decade 1 to 10 mH.
R_w	Decades 0·1 to 100 ohms.
C_E	Decades 0·1 to 10 μF.
T_1 and T_2	Test terminals.
S_1	Ratio/substitution switch.
S_2	Standard capacitor selector switch.
S_3	Bridge to Wagner switch.
T_x	Step-up transformer with mumetal screen.

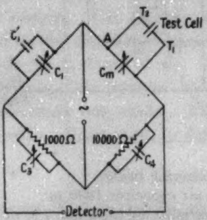
Broken lines denote shields.

General experience shows that the accuracy obtained is usually greatest for capacitances of 500 pF and upwards, and since the capacitance of the empty cell was only 60 pF, it became necessary to abandon the substitution method of measurement employed by Hartshorn and Oliver (1929) and to use the ratio bridge shown in figure 3, which is realized by switching the cell connected to T_1T_2, and the precision trimmer capacitor C_m in parallel with it, into the arm 2 of the bridge, and connecting the ballast capacitor C_2 shown in figure 1 to earth E. The trimmer capacitor C_m allows the initial reading of the standard capacitors

C_1 obtained with the cell disconnected to be brought to a convenient, accurately reproducible setting, near the bottom of the scale. Then on connecting the cell, the second reading at balance was some 600 pF higher and therefore near the top of the scale of the standard capacitor of total value 750 pF. The observations were repeated for several settings of the trimmer capacitor, to ensure that observational errors and calibration errors were insignificant.

The cell is connected to the terminals T_1T_2 by lengths of screened coaxial cable as already mentioned. The disconnection could be made with both leads or one only, and either at the end of the cable attached to the bridge or at the end attached to the cell. Provided the screening is adequate all these methods of disconnection give the same reading of capacitance, the settings of the Wagner arms alone requiring adjustment. The adequacy of the screening was verified by observing these four readings for every set-up of the bridge used.

The capacitance of the cell when filled with benzene was about 140 pF, so that a difference of capacitance of 1400 μpF was required in arm 1 to balance it. This was obtained by using the fixed capacitor of 1000 pF in parallel with the variable standard employed for the measurements on the empty cell.



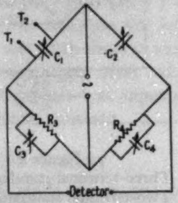


Figure 3. Schematic arrangement of Schering bridge for measurements on benzene and cyclohexane.

C_1 Standard variable air capacitor 750 pF.
C_1' Standard air capacitor 1000 pF.
C_m Variable air capacitor 2 to 6 pF.

Figure 4. Schematic arrangement of Schering bridge for measurements on polar liquids.

C_1 Standard variable air capacitor 750 pF or 150 pF.
C_2 Variable air capacitor 100 to 1000 pF. $R_3 = R_4 = 1000\ \Omega$.

The difference of capacitance ΔC_v required to balance the evacuated cell, and the corresponding difference ΔC_l required to balance the cell when filled with liquid, give the dielectric constant ϵ_l of the liquid, the simple relation $\epsilon_l = \Delta C_l / \Delta C_v$ being free from corrections, provided that the ratio arms 3 and 4 of the bridge remain unchanged throughout both series of observations. Actually, the finite conductivity of the liquid necessitates a slight adjustment to C_3 or C_4 in changing from the condition cell empty to the condition cell full, but the resulting correction only amounts to 1 part in 10^4 when the loss tangent of the liquid is as great as 0·01. The loss tangent for the pure benzene was found to be never more than 0·001, so that no correction was required, and the accuracy of the measured dielectric constant depends only on the accuracy of measurement of the capacitance differences and the stability of the cell and the liquid.

The measurements of loss tangent are best made by substitution, a variable air capacitor being employed as a standard of constant power loss, independent of the capacitance setting. These measurements were therefore made by switching the cell into arm 1 of the bridge, removing the trimmer capacitor, and connecting a ballast capacitor into arm 2 as shown in figure 4. Such measurements showed that the loss tangent of the benzene was completely determined by its water content. The value was only measurable at low frequencies, and varied from 0·001 at 1 kc/s for an undried sample to 0·00005 for a dried sample. As such values seem to have no intrinsic interest they were not recorded in detail. At the higher frequencies the values were too small to be detected with certainty, though when the frequency reaches the microwave region larger values are again obtained. These are discussed in Part III.

§ 3. Results for Benzene

Samples of specially purified benzene were supplied by Professor J. Timmermans from the Union Internationale de Chemie at Brussels, and by Dr. Herington from the Chemical Research Laboratory, Teddington. For purposes of comparison with material generally available measurements were also made on two samples of benzene of analytical reagent grade supplied by two makers.

The samples were first distilled into the cell under vacuum without previous treatment and the dielectric constant measured. They were then dried by standing over powdered calcium hydride and the measurements of dielectric constant were repeated after various periods of drying. As was expected, there was a diminution of dielectric constant as drying proceeded but after a period of three months each sample appeared to have reached a stationary condition. They were then refluxed under vacuum in the presence of calcium hydride for one to two weeks, a process calculated to accelerate any further drying that might be possible, but no further change in dielectric constant was observed. It is therefore concluded that the stationary value obtained is the dielectric constant of the perfectly dry liquid.

Measurements were made at frequencies covering the range 50 c/s to 10 kc/s but the dielectric constant was found to be independent of frequency, and the values given in table 3 may therefore be regarded as 'static' values, holding good for all frequencies below a certain limit, which is shown by these measurements to be higher than 10 kc/s.

It is to be noted that there were comparatively large differences between the samples as first received, but that these were almost entirely due to differences of water content. The differences between the values for Professor Timmermans' and Dr. Herington's samples in the dry state were within the experimental error at all temperatures and it is therefore possible to state as the final result, for the dielectric constant of pure dry benzene at 20·00°c (ϵ_{20}), and for the value ϵ_t at any temperature within the range 20° to 30°c,

$$\epsilon_{20} = 2{\cdot}2836 \pm 0{\cdot}0002$$

$$\epsilon_t = \epsilon_{20} - 0{\cdot}0019\,(t-20).$$

The limits of error shown are mainly determined by the stability of the standard capacitors and the cell.

The values of dielectric constant and temperature coefficient given in table 3 are averages, some six or eight readings being taken for each value of capacitance. The standard deviation for the values of dielectric constant was about 0·0001.

The temperature coefficient of dielectric constant is shown by table 3 to be considerably affected by impurities. Unfortunately the increase produced by water is apparently liable to be compensated by a diminution produced by other impurities, so that the actual value for any sample cannot be safely used as a criterion of purity.

Table 3. Benzene (Temperature 20°c)

Sample and Treatment	Dielectric Constant (Vacuum = 1)	Temperature Coefficient *a*
Prof. Timmermans		
1st Distillation	2·2852	0·0019
1 week drying	2·2842	0·0019
4 weeks drying	2·2840	0·0019
3 months drying	2·2836	0·0019
3 months drying and 2 weeks refluxing	2·2836	0·0019
Chemical Research Laboratory		
1st Distillation	2·2867	0·00195
1 week drying	2·2848	0·0019
4 weeks drying	2·2837	0·0019
3 months drying	2·2835	0·0019
3 months drying and 2 weeks refluxing	2·2835	0·0019
Hopkin and Williams		
As received	2·2884	0·0020
1st Distillation	2·2863	0·0019
1 week drying	2·2843	0·00185
3 weeks drying	2·2836	0·00185
3 months drying	2·2836	0·00185
3 months drying and 1 week refluxing	2·2836	0·00185
British Drug Houses		
As received	2·2890	0·0020
1st Distillation	2·2858	0·0019
1 week drying	2·2850	0·00185
2 weeks drying	2·2840	0·00185
3 months drying	2·2838	0·00185

§ 4. CYCLOHEXANE AS A STANDARD

The difficulty of removing water from benzene is a serious objection to its use as a standard liquid for dielectric measurement. Previous measurements of the audio-frequency power factor of commercial grade cyclohexane indicated that the water content was exceptionally low, and consequently a sample of pure cyclohexane was obtained from Professor Timmermans and was given the same treatment as the benzene samples. The dielectric constant was measured before treatment and this value was found to be completely unaffected by the drying procedure. The values obtained for cyclohexane were

$$\epsilon_{20} = 2{\cdot}0250 \pm 0{\cdot}0002$$

$$\epsilon_t = \epsilon_{20} - 0{\cdot}00155(t-20)\ldots \text{for } 20^\circ\text{c} < t < 30^\circ\text{c}.$$

It was evident from the results obtained that for use as a standard liquid cyclohexane has decided advantages over benzene.

Cyclohexane unfortunately is not available in an analytical reagent grade. The dielectric constant of a commercial sample was measured and was found to be 2·0300.

§ 5. METHOD FOR POLAR LIQUIDS

On proceeding to make measurements on dichlorethane difficulties of a new kind were encountered. A preliminary measurement on a small sample showed that its conductivity was so high that its loss tangent at 1 kc/s amounted to 0·5. In such circumstances measurements of the precision obtained with benzene were out of the question. The three-terminal cell of figure 1 was therefore abandoned for all the preliminary work on polar liquids in favour of a simpler two-terminal cell designed some years ago for general work on liquids. It is shown in figure 5. The electrodes are cylinders machined from Invar and rhodium plated and the outer cylinder acts as container for the liquid. A conical bushing of fused quartz provides almost perfect insulation and precise location for the inner cylinder. The chief advantages of this cell for the work in hand are the ease of dismantling and cleaning, and the small quantity of liquid required, nearly the whole of the liquid being subjected to the full value of the applied electric field. Inner electrodes of various diameters are available. For dichlorethane one giving a gap of 1 mm and requiring only 15 ml. of liquid was convenient; for nitrobenzene a gap of 2 mm and 25 ml. of liquid were used.

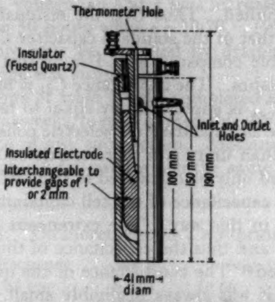


Figure 5. Two-terminal cell.

With this simpler cell the capacitance measured includes that located in the quartz insulator as well as that in the liquid or the empty space which it subsequently occupies, and this capacitance amounts to 20% of the total capacitance of the empty cell. It follows that measurements of dielectric constant cannot be made with the accuracy obtainable with the cell of figure 1. However, by using benzene previously measured in the latter cell as a standard it is possible to make measurements in the simpler cell with all the accuracy that seemed likely to be significant when dealing with liquids of such high conductivity as the dichlorethane. The procedure is to use a constant volume of liquid and to assume that the capacitance extraneous to this volume is constant. Since the lines of electric force in the cell lie in the liquid surface, this condition is fulfilled to an accuracy determined by the stability of the cell on filling and emptying.

If C_1 is the measured capacitance of the cell when charged with the fixed volume of liquid, C_v is the measured capacitance with the cell evacuated, and C_e is the capacitance extraneous to the fixed volume, then $\epsilon=(C_1-C_e)/(C_v-C_e)$. The measurement on benzene served to determine C_e for the particular volume of liquid to be used and the value of ϵ for other liquids was then determined by applying the same equation.

Since it became clear at the outset that the measurements on polar liquids would be dominated by considerations of conductivity or power loss, the measurements were made by the bridge arrangement of figure 4, that is to say by substitution for the difference of capacitance between two settings of the standard air capacitor C_1 in a bridge with equal ratio arms. The simplicity of this arrangement and its comparative freedom from troublesome corrections are advantageous in measurements in which the loss tangent is an important factor. The procedure is to balance the bridge for the two conditions: (*a*) with the cell connected to T_1T_2 and (*b*) with the cell disconnected, the capacitors C_1 and C_4 alone being varied. If $C_{1a}\,C_{4a}$ are the two readings of these capacitors with cell connected, and $C_{1b}\,C_{4b}$ the readings with the cell disconnected, then the capacitance of the cell C_c and its conductance G_c are given by

$$C_c(1+\omega^2R_4^2C_{4a}^2)=(C_{1b}-C_{1a})-G_{1b}R_4(C_{4a}-C_{4b})$$
$$-\,\omega^2R_4^2C_{4a}^2\left[C_{1a}-\frac{C_{4b}}{C_{4a}}C_{1b}\right] \qquad \ldots\ldots(1)$$
$$G_c=\omega^2R_4[(C_{1a}+C_c)C_{4a}-C_{1b}C_{4b}]. \qquad \ldots\ldots(2)$$

No approximation has been made in writing down the above equations; in practice they can usually be simplified. Thus G_{1b}, the residual conductance in arm 1 of the bridge, is mainly that of the standard capacitor C_1. It can be measured and shown to be negligible compared with $1/R_4$ so that the middle term on the right-hand side of (1) vanishes. The capacitance C_{4b} is made as small as possible and then the factor ωR_4C_{4a} is approximately equal to $\tan\delta$ of the liquid. For any liquid suitable for use as a standard of dielectric constant this quantity must be considerably smaller than unity and therefore its square is smaller still and the terms including it need only be calculated to two or three significant figures.

The conductance and capacitance of the cell containing the standard volume of liquid were measure[illegible]this way. The extraneous capacitance mentioned above was then deducted and thus the capacitance of the liquid C_s and its conductance G_s were obtained. The conductance of the quartz bushing could be separately measured but it was always negligibly small. For convenience the loss tangent $\tan\delta_s=G_s/C_s\omega$ is recorded instead of G.

§ 6. EXPERIMENTS ON DICHLORETHANE

The sample of specially purified dichlorethane supplied by Professor Timmermans had been sealed off under vacuum, and although the seal had not been broken there was a delay of some months before the measurements could be made. The first measurements showed that the liquid possessed a conductivity of the order of 3×10^{-9} mho cm^{-1} ($\tan\delta=0{\cdot}5$ at 1 kc/s) which suggested the presence of a considerable concentration of free ions, which could only be regarded as an impurity produced by decomposition of the pure liquid during storage. A colleague, Mr. E. Rushton, suggested that the free ions could be conveniently removed by applying to the cell a d.c. voltage; it is merely necessary to connect a battery in series with the a.c. generator used for the measurements of capacitance. Such a d.c. voltage has no observable effect on the measured values for liquids of negligible conductivity but the effects observed with our polar liquids were considerable.

Table 4 gives the results obtained with the sample of dichlorethane in the two-terminal cell, the applied d.c. voltage being 300 v and the electrode spacing

1 mm. The values were independent of the a.c. voltage so long as it was small compared with the d.c. voltage.

Table 4. Dichlorethane

	$f=1$ kc/s		$f=10$ kc/s	
t	ϵ_{20}	tan δ	ϵ_{20}	tan δ
0	10·85	0·50	10·75	0·049
2 min		0·33		
10 min	10·78	0·20		
3 hours		0·071		
24 hours	10·66	0·034		
48 hours	10·66	0·008	10·66	0·0008

t = time interval between the application of the d.c. voltage and the making of the measurement.

ϵ_{20} = dielectric constant at 20·00°c.

These measurements suggest that the stationary value of dielectric constant that is attained when tan δ is reduced below, say 0·03, is characteristic of the pure liquid.

On removing the d.c. voltage tan δ at 1 kc/s increased from 0·008 to 0·08 in 20 minutes and to 0·20 in 50 hours when $\epsilon=10{\cdot}70$. Permanent removal of the impurities could therefore only be obtained by draining the liquid from the electrodes while the voltage was still applied.

After the measurements of table 4 were completed the liquid was kept over calcium hydride for several weeks, to see if appreciable drying would occur. There was however no evidence of any such action. Measurements made under the same conditions as for table 4 followed much the same course, the final results being shown in table 5.

Table 5. Dichlorethane in Two-terminal Cell

Frequency	tan δ	ϵ_{20}
65 c/s	0·104	10·66
1 kc/s	0·008	10·66 ± 0·01
10 kc/s	0·0008	10·66 ± 0·01

For $20°\text{c} > t > 30°\text{c}$, $\epsilon_t = \epsilon_{20} - 0{\cdot}0555\,(t-20)$.

The limits of error on the value of dielectric constant were mainly determined by the limitations of the two-terminal cell. An attempt was therefore made to improve on this value by the use of the three-terminal cell. However, although this successfully eliminates the main errors associated with capacitance in the quartz or glass insulators it is less satisfactory for the process of removing ions by application of d.c. voltage, for only part of the sample lies in the electric field between the cylinders and therefore ions continually diffuse into the field from the external liquid, so that the process is necessarily slower and less complete. However, after maintaining the dichlorethane in the cell under a voltage of 300 v for 3 months, the value of tan δ had reached the stationary values shown in table 6. There is therefore good reason to believe that the corresponding value of dielectric constant is the static value for the pure liquid. It is very satisfactory to note that the value obtained with the two-terminal cell is confirmed within its estimated error and that the use of the three-terminal cell reduces the uncertainty to about a third of its former value.

The accuracy obtained in this case is necessarily somewhat inferior to that obtained with the hydrocarbons because the liquid itself is less stable, and the balance point of the bridge could not be determined with the same precision. The standard deviation for the observations, which when averaged gave the values of ϵ_{20} in table 6, was approximately 0·002.

Table 6. Dichlorethane in Three-terminal Cell

Frequency	tan δ	ϵ_{20}
1 kc/s	0·031	10·663 ± 0·003
10 kc/s	0·0033	10·663 ± 0·003

For 20°c > t > 30°c, $\epsilon_t = \epsilon_0 - 0{\cdot}553\ (t-20)$.

§ 7. MEASUREMENTS ON NITROBENZENE

The decomposition on standing seemed even more marked with nitrobenzene than with dichlorethane. The sample of initially pure mono-nitrobenzene supplied by Professor Timmermans, when poured into the two-terminal cell, was found to have so high a value of tan δ, even after a prolonged application of a suitable d.c. voltage, that no satisfactory measurement of its dielectric constant could be made at audio frequencies. The liquid was therefore distilled; about a third of the distillate, boiling at 203° to 209°c, was rejected and the main fraction boiling at 211°c was poured into the two-terminal cell, maintained under a d.c. voltage gradient of 150 v mm^{-1} for 48 hours, and then measured. The values obtained are given in table 7.

Table 7. Nitrobenzene (1st Sample)

Frequency	tan δ	ϵ_{20}
100 c/s	2·0	36
1 kc/s	0·190	35·75 ± 0·02
10 kc/s	0·018	35·75 ± 0·02

For 20°c > t > 30°c, $\epsilon_t = \epsilon_{20} - 0{\cdot}185\ (t-20)$.

The electrodes of the cell, after treatment of the nitrobenzene with the d.c. voltage, were found to be coated with a blackish deposit, which indicated that some impurity had been removed from the liquid. It should however be mentioned that the quantity was not sufficient to affect the capacitance of the cell when empty by an appreciable amount, so that it is not likely to have directly affected the values obtained for the dielectric constant. Nevertheless further purification seemed desirable. Unfortunately the sample was not at this stage large enough to permit further rejection and since the avoidance of any delay between purification and measurement seemed to be of the first importance, it was decided to prepare a fresh sample beginning with commercially pure nitrobenzene (analytical reagent grade). After repeated distillation, in each case rejecting about one-third of the distillate as before, the final distillate was placed under a d.c. voltage gradient of 150 v mm^{-1} for three days, and the liquid then removed from the electrodes while the voltage was still applied. This treatment presumably brought it to the condition of the first sample, or nearly so. It was then refluxed over finely powdered alumina for 24 hours, a process which, judged by the greatly changed colour of the alumina, removed further impurities. It was finally distilled directly into the two-terminal cell, the d.c. voltage applied as before and a measurement made. The results are given in table 8.

The liquid was afterwards allowed to stand over powdered calcium hydride for four weeks and then re-distilled into the cell and the measurements repeated. The final results were identical with those of table 8. The values of tan δ leave no room for doubt that the final sample was purer than that of table 7 and since the value of dielectric constant in table 8 is clearly reproducible and independent of frequency, it may be taken for all practical purposes as the static dielectric constant of pure nitrobenzene. It is possible that the limits of error of the final value could be slightly reduced by the use of the three-terminal cell, but the additional labour could scarcely be justified unless a more stable liquid could be found for use as a standard.

Table 8. Nitrobenzene (Final Sample)

Frequency	tan δ	ϵ_{20}
100 c/s	0·15	35·7
1 kc/s	0·0175	35·72 ± 0·02
10 kc/s	0·0022	35·72 ± 0·02

$$\epsilon_t = \epsilon_{20} - 0{\cdot}185\ (t-20).$$

PART III. MEASUREMENTS AT MICROWAVE FREQUENCIES

By L. Essen and J. V. L. Parry

§ 1. Scope of the Measurements

The dielectric constants measured in Part II were shown to be static values, which hold good for all frequencies below the range in which an absorption band first appears when the frequency is continuously increased. In any such range a diminution of dielectric constant with rise of frequency is to be expected. No absorption band, or peak in the curve of tan δ plotted against frequency, has been observed for any of our standard liquids at frequencies below the microwave range and it is believed therefore that the values given in Parts I and II are valid for all radio as well as audio frequencies. When the microwave region is reached the position is less certain.

Investigations in the microwave region of the liquids measured by us have been carried out by Jackson and Powles (1946), Penrose (1946), Bleaney, Loubser and Penrose (1947) and Whiffen (1946, 1950) among others, and it is clearly established that in nitrobenzene strong absorption and dispersion occur at about 10 000 Mc/s. In the hydrocarbons absorption remains relatively small even at frequencies up to 24 000 Mc/s, but at this frequency it is increasing rapidly with frequency, the increase being much greater for benzene than for cyclohexane. The increase is not so large as to lead one to expect to find measurable dispersion, even for benzene, below 24 000 Mc/s, and Bleaney, Loubser and Penrose (1947), who alone have reported direct measurements of dielectric constant in this region, could detect no change of value between 9000 and 25 000 Mc/s. On the other hand, although their value for cyclohexane agrees with our own value, their value for benzene is lower than ours by nearly 0·3%, which is nearly ten times their estimated experimental error. It is difficult to account for this low value by anything except dispersion. In the face of this discrepancy it was obviously important to make a new direct determination of the dielectric constant of benzene at 9000 Mc/s.

It seemed also desirable to make a sufficient number of measurements of tan δ on benzene, cyclohexane and on dilute solutions of nitrobenzene in benzene to show whether the values of this quantity for our specially purified materials were appreciably different from those previously published.

§ 2. THE DIELECTRIC CONSTANT OF BENZENE AT 9000 Mc/s

2.1. *Method*

The work of Essen and Froome (1951) has shown that the cavity resonator method is capable of giving very high precision when applied to gases. In the present work a similar technique was applied to liquids but it soon became apparent that the use of a liquid imposes a considerable restriction on the accuracy, which depends primarily on the reproducibility of the liquid-filled resonator and its sharpness of resonance Q. A cavity designed to contain liquids of the highest purity must be kept small since a large volume of liquid is never available, and this condition limits the value of Q obtainable. An additional limitation is imposed by power loss in the liquid itself. It will be shown in a later paragraph that tan δ for pure benzene is 0·0005 at 9000 Mc/s and increases with further rise of frequency. It follows that Q for a benzene-filled cavity cannot exceed 2000 at such frequencies even for the most favourable design. However, even with this restriction it seemed that an accuracy comparable with that obtained at low frequencies should be obtainable without undue elaboration.

The cavity-resonance method depends on the basic relation between the frequency f, the length L and diameter D of a uniform cylindrical cavity and the absolute permittivity ϵ_a and permeability μ_a of the electromagnetic medium filling it. This relation may be written

$$\epsilon_a = \frac{c^2}{\mu_a}\frac{1}{f^2}\left[\left(\frac{\Gamma_{lm}}{\pi D}\right)^2 + \left(\frac{n}{2L}\right)^2\right] \qquad \ldots\ldots(3)$$

where lmn are integers characteristic of the mode of vibration. For H modes, which were used in this work, Γ_{lm} is the mth root of $J_l(x)$ and n is the number of half-wavelengths in the length L of the cavity.

This relation is only strictly valid for an ideal cylindrical cavity totally enclosed by walls of zero resistivity. In practice the walls must be constructed of material of finite resistivity and they cannot be completely closed since holes must be provided for the emptying and filling of the cavity and for coupling to the wave-guide of the measuring apparatus. Thus when applied to a real cavity, equation (3) can only be used as an approximation, but it does provide a means of measuring the permittivity ϵ_a of a medium in terms of the general constant c^2/μ_a, which has the same value for all non-magnetic media, and the quantities f, D, L, l, m, n, all of which can be determined experimentally with very high precision. The accuracy of such a measurement of ϵ_a is therefore limited mainly by the closeness of the approximation mentioned above, and in order to improve it, it would be necessary to include in the equation terms representing the effects of departures from the ideal cavity, namely, the holes, etc. which in our cavities were made in an end wall (' end effects ') and the penetration of the electromagnetic field into the walls (' skin effect '). These effects may be regarded as imposing corrections on the values of D and L as measured by mechanical devices.

The effects of these sources of error on a measurement of dielectric constant can be greatly reduced by making two measurements of resonant frequency f on the same cavity, one with the cavity filled with the liquid and one with it evacuated. It is possible to design a cavity so that for a particular liquid these two frequencies have the same value, for some pair of values of n_l the number of half-wavelengths in the length L of liquid and n_v the number of half-wavelengths in the length of L the vacuum, the choice of these numbers depending on the dielectric constant of the liquid, which has been given approximately by (3). The dielectric constant or relative permittivity ϵ of the liquid is then given with a closer approximation by taking the ratio of the values of ϵ_a for the liquid and the vacuum determined by the two sets of observations. Thus

$$\epsilon=\left(\frac{\Gamma_{lm}}{\pi D}\right)^2+\left(\frac{n_l}{2L}\right)^2\Big/\left(\frac{\Gamma_{lm}}{\pi D}\right)^2+\left(\frac{n_v}{2L}\right)^2. \qquad \ldots\ldots(4)$$

Since the observations are all made at the same frequency the correction terms associated with D and L will be nearly the same in both cases and on taking the ratio they will largely cancel one another.

The residual error can in principle be estimated and corrected by repeating the whole series of measurements with another cavity differing in length only from the first one, and thus resonating at the required frequency with different values of n. The apparent change in ϵ caused by a measured change in L provides an indication of the magnitude of the error which can therefore be corrected.

In this connection the use of a single cavity with a movable piston is attractive as permitting a determination of $\Delta L=\lambda/2$ free from end effects, but it is not readily practicable in a vacuum-tight form in which, moreover, no grease or rubber is permissible. Two cavities differing in length only were therefore tried, but as will appear later the residual error was masked by the accidental errors associated with measurements on a liquid-filled cavity in the present state of the art.

2.2. *Experimental Details*

The cavities were designed to resonate at a frequency of 9200 Mc/s because a very stable oscillator at that frequency was already available. They resonated in the H_{011} and H_{012} modes when evacuated and in the H_{012} and H_{014} modes when filled with benzene. Their dimensions were as follows: small cavity, internal diameter 5·2522 cm, length 2·4923 cm, large cavity, internal diameter 5·2524 cm, length 4·9842 cm, wall thickness 1·25 cm.

The material used was Invar with silver plating followed by a rhodium flash on the inside surfaces. The end faces of the cylinders and the end plates were lapped and optically polished. The same end plates were used for the two cavities so that the end effect would be the same in the two cases. In one plate a narrow channel was machined having an outside diameter equal to the internal diameter of the cylinder, in order to move the frequency of the degenerate E_{11} modes away from that of the H_{01} modes. It became evident during the observations that this was not completely effective and in addition a mica washer 0·004 cm thick was therefore placed between the barrel of the cavity and the end plate. The thickness of this was taken into account in the measurements by deducing L from the overall length and the thickness of the end plates. The other end plate was provided with a coupling hole sealed with polythene, leading through a waveguide to the oscillator and with a hole leading through copper bellows and a glass–metal seal to an all-glass distillation apparatus. The coupling hole was adjusted in size to match the cavity at resonance to the waveguide and the other hole was smaller and situated near the circumference of the cavity, where the field is weak, so as to have a negligible effect on frequency.

The arrangements for filling the cavity with pure benzene by low-temperature distillation were the same as were used for Part II.

The cavity under investigation was immersed in an oil bath which could be maintained within ±0·01°c of the required temperature. The resonant frequencies were measured, in exactly the same way as that used by Essen and Froome (1951). As shown in figure 6, the liquid measuring cavity was connected

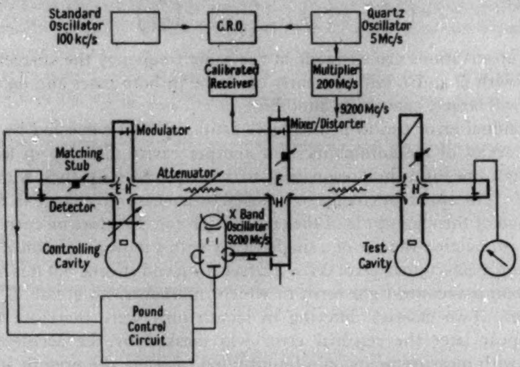

Standard Oscillator 100 kc/s
C.R.O.
Quartz Oscillator 5 Mc/s
Calibrated Receiver
Multiplier 200 Mc/s
9200 Mc/s
Mixer/Distorter
Modulator
Matching Stub
Attenuator
E
H
E
H
H
H
Detector
X Band Oscillator 9200 Mc/s
Controlling Cavity
Test Cavity
Pound Control Circuit

Figure 6. Cavity bridge and frequency comparator.

in one arm of a waveguide bridge which is balanced at resonance. As the frequency of the source departs from the resonance value a large out-of-balance signal is detected and the bridge thus provides a very sensitive means of setting to resonance The frequency is measured directly in terms of harmonics from the quartz standard with a precision which can easily be made to exceed 1 part in 10^6.

2.3. *Experimental Procedure and Results*

The cavities which had purposely been left slightly longer than theoretically required were ground down until the resonances were at a convenient frequency and were approximately the same for the evacuated and filled condition. The resonant frequency of the evacuated cavity was then measured more carefully and the cavity was filled by distillation with benzene of analytical reagent grade, and re-measured. This cycle was repeated a number of times to check the repetition accuracy of the experiment. The results obtained with the filled cavity were far less constant then expected. Although setting to resonance by means of varying frequency is practicable and accurate under suitable conditions some means of tuning the cavity would have been very useful in initial tests made to establish these conditions. Nevertheless the accuracy was comparable with that of the low-frequency measurements and was therefore adequate for our immediate purpose.

Measurements were made with both cavities using the same benzene and the results obtained using equation (2) were the same within the experimental accuracy. The second cavity therefore served the purpose of showing that the residual errors obtained by using this equation were not of significance and that there was no need to make measurements with both cavities in every case. Measurements were finally made with the large cavity only using pure dried benzene. The results are summarized in table 9.

Table 9. The Dielectric Constant of Benzene at 9200 Mc/s and 20°c

Analar Benzene (1 distillation)	Small cavity 2·2862
Analar Benzene (1 distillation)	Large cavity 2·2864
Pure Benzene, dried	Large cavity 2·2841

Estimated limits of error ±0·0005

The results agree within the experimental accuracy with those given in table 3 and show that there is no detectable dispersion between audio frequencies and 9200 Mc/s. The value of the temperature coefficient was also measured and was found to agree with the low-frequency value.

It will be observed that the accuracy obtained at 9200 Mc/s was not quite so good as that at audio frequencies, and that somewhat surprisingly it was mainly limited by lack of reproducibility of the resonant frequency of the liquid-filled cavity. It was by no means certain that the full accuracy of which the method is capable had been reached but, since the effort required for a detailed investigation of the uncertainties seemed out of proportion to the probable gain, the method was not pursued further.

§ 3. MEASUREMENTS OF POWER LOSS

Measurements of power loss were made on small samples inserted in a cavity of adjustable length vibrating in the H_{01} mode, using the method described by Lamb *et al.* (1946). The results for benzene and cyclohexane are given in table 10.

Table 10

		tan δ	
		9000 Mc/s	25000 Mc/s
Benzene, A.R.	undried	0·0005	0·0012
	dry	0·0005	0·0009
Benzene, C.R.L.	undried	0·0005	0·0011
	dry	0·0004	0·0009
Cyclohexane	undried	0·0002	0·00025
	dry	0·0002	0·0002

The values are in good agreement with those of Whiffen (1950) and support his conclusion that the loss measured in dry benzene is associated with the benzene molecule itself.

The values for a dilute solution of nitrobenzene in benzene are given in table 11, in which the values of tan δ have been reduced to those for a concentration of 1 mole per litre to facilitate comparison with the earlier results collected by Whiffen (1950).

When these results are inserted in the Debye equation in the form used by Whiffen it is found that they correspond to a dipole moment for nitrobenzene in benzene solution of 4·05 D and a time of relaxation of $11{\cdot}4 \times 10^{-12}$ sec, and moreover that the accuracy with which these quantities are determined is rather better than 1%. Whiffen's own latest values were 3·92 D and $12{\cdot}8 \times 10^{-12}$ sec, but earlier values determined by Whiffen himself and by Jackson and Powles (1946) agree so well with our own that we think they must be the better values, and that the differences arise from the methods of measurement rather than differences in purity of material.

Table 11

Frequency (Mc/s)	2950	9200	24000
Tan δ (molar)	0·184	0·419	0·400

ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

BLEANEY, B., LOUBSER, J. H. N., and PENROSE, R. P., 1947, *Proc. Phys. Soc.*, **59**, 185.
ESSEN, L., and FROOME, K. D., 1951, *Proc. Phys. Soc.* B, **64**, 862.
HARTSHORN, L., and OLIVER, D. A., 1929, *Proc. Roy. Soc.* A, **123**, 664.
JACKSON, WILLIS, and POWLES, J. G., 1946, *Trans. Faraday Soc.*, **42** A, 101.
LAMB, J., HORNER, F., TAYLOR, T., DUNSMUIR, R., and JACKSON, WILLIS, 1946, *J. Instn Elect. Engrs*, **93**, Part III, 53.
MARYOTT, A. A., and SMITH, E. R., 1951, *Table of Dielectric Constants of Pure Liquids, Circ. U.S. Bur. Stand.* 514.
PENROSE, R. P., 1946, *Trans. Faraday Soc.*, **42** A, 155.
POWLES, J. G., 1946, *Trans. Faraday Soc.*, **42** A, 157.
WHIFFEN, D. H., 1950, *Trans. Faraday Soc.*, **46**, 124.
WHIFFEN, D. H., and THOMPSON, H. W., 1946, *Trans. Faraday Soc.*, **42** A, 114.

A Pulse Method for Measuring the Injection Ratio of Metal–Semiconductor Contacts

By G. G. E. LOW
Department of Physics, University of Reading

Communicated by R. W. Ditchburn ; *MS. received 2nd February* 1955

Abstract. A method for measuring the injection ratio of metal–semiconductor contacts is described. It is suitable for measurements at low emitter currents. The current through the contact is applied as a series of constant current pulses and the change in conductance of the specimen is measured as a function of time. Measurements have been carried out for tungsten point contacts on n-type and p-type germanium and results of injection ratio as a function of emitter current are given.

---

§ 1. INTRODUCTION

THE current through a metal–semiconductor contact is, in general, carried by both holes and electrons and the injection ratio γ is defined as the fraction of the total current carried by minority carriers. For near intrinsic semiconductors it is useful to define a second ratio γ_0 as the fraction of the total current carried by minority carriers in the bulk of the material when the hole and electron concentrations have their equilibrium values. Many experiments are concerned with disturbances in the minority carrier concentration and consequently they depend on a quantity given by $\gamma_e = \gamma - \gamma_0$. It is this quantity which is determined by the methods for measuring injection ratio previously reported and also by the method described in the present paper.

Methods for measuring the injection ratio have been described by Shockley, Pearson and Haynes (1949) and by Many (1954). The method of Shockley, Pearson and Haynes is satisfactory for contacts of high injection ratio, but it involves an extrapolation to infinite sweeping field. It is thus assumed that the injection ratio of the contact is constant for the range of sweeping fields used in the experiment. The method here described does not involve this extrapolation. Moreover point contact voltage probes are not used and each determination involves relatively few readings. The Many bridge circuit provides a convenient method for measurements on contacts of very low injection ratio. Although in some respects similar to Many's method, the present experiment avoids this limitation by separating the emitter and auxiliary contacts. The sensitivity of the experiment is such that the injection ratio of a contact can be measured at small emitter currents. Measurements have been carried out for tungsten point contacts on n-type and p-type germanium.

§ 2. THEORY OF EXPERIMENTAL METHOD

There are two variants of the present method. In both cases it is assumed that the disturbance in minority carrier concentration is small compared with the equilibrium concentration of majority carriers; that, in the longitudinal direction, diffusion currents are negligible compared with those due to the electric field; and that the injected carriers all decay in the filament and are not swept out through the end contact. These assumptions are involved in the expression for the change of conductance derived by Many (1954), which is used below.

2.1. *Emitter Current applied in Square Pulses*

The circuit used is shown in figure 1. For the present method the battery E_2 is removed. The filamentary specimen (assumed to be n-type in the present case) with low resistance end contacts is connected so as to form two arms of a bridge circuit. The position of the contact under examination divides the specimen into two parts of lengths L_1 and L_2, and having initial resistances R_1 and R_2 respectively. The network is initially balanced in the absence of a sweeping current, i.e. with E_1 zero, so that $R_1R_4 = R_2R_3$. The pulse generator can be used for this purpose. The battery E_1 is then inserted resulting in a sweeping current I_s. For zero emitter current

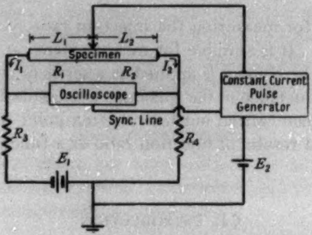


Figure 1. Circuit of pulse method for measuring the injection ratio of contacts.

$$I_1(R_1+R_3) - I_2(R_2+R_4) = E_1 \qquad \ldots\ldots(1)$$

and

$$I_1 = -I_2 = I_s = E_1/(R_1+R_2+R_3+R_4). \qquad \ldots\ldots(2)$$

The values of R_3 and R_4 are chosen sufficiently large so that the sweeping current I_s is practically independent of small changes in filament resistance. When the constant current pulse is applied (positive for n-type material), a current δI_e flows through the contact under test. This results in I_1 changing to I_1' and I_2 to I_2'. If the resistances R_1 and R_2 were unchanged the bridge circuit would remain balanced and no voltage change between the ends of the filament would be recorded by the oscilloscope. However, if the test contact is injecting, R_2 remains unaffected but R_1 changes to R_1', since the injected holes drift towards the negative end of the filament. The change $R_1 - R_1'$ is, however, so small as to leave the emitter current constant for practical purposes. Under these conditions

$$I_1'(R_1'+R_3) - I_2'(R_2+R_4) = E_1 \qquad \ldots\ldots(3)$$

where

$$I_1' + I_2' = \delta I_e. \qquad \ldots\ldots(4)$$

The voltage change across the ends of the filament is therefore given by

$$\begin{aligned}\delta V &= R_3(I_1'-I_1) - R_4(I_2'-I_2)\\ &= (R_3+R_4)(I_1'-I_1) - R_4\delta I_e \qquad \ldots\ldots(5)\end{aligned}$$

(using equation (4)). From equations (3) and (4) it can be shown that I_1' consists of two parts, namely

$$I_1' = \frac{E_1}{R_1' + R_2 + R_3 + R_4} + \frac{\delta I_e(R_2 + R_4)}{R_1' + R_2 + R_3 + R_4} \qquad \ldots\ldots(6)$$

of which the first term represents the modified circulating current and the second the appropriate fraction of the injected current. Combining equations (1), (2), (5) and (6) leads to

$$\delta V = \frac{E_1(R_3 + R_4)(R_1 - R_1')}{(R_1 + R_2 + R_3 + R_4)(R_1' + R_2 + R_3 + R_4)} + \frac{\delta I_e R_4(R_1 - R_1')}{R_1' + R_2 + R_3 + R_4}. \qquad \ldots\ldots(7)$$

The difference between R_1 and R_1' is small in accordance with the assumptions set out earlier. The last expression may thus be simplified to

$$\delta V = \frac{I_s(R_3 + R_4) + \delta I_e R_4}{R_1 + R_2 + R_3 + R_4} R_1^2 \delta G_1 \qquad \ldots\ldots(8)$$

where δG_1 is the change in conductance given by

$$\delta G_1 = -R_1^{-2} \delta R_1 = -R_1^{-2}(R_1' - R_1). \qquad \ldots\ldots(9)$$

Many (1954) has shown that the change in conductance of a filament resulting from the application of a constant current pulse to an emitter contact is given by

$$\delta G_1 = \mu_p(b+1) L_1^{-2} \gamma_e \delta I_e \tau \{1 - \exp(-t/\tau)\} \qquad \ldots\ldots(10)$$

where t is the time elapsed since the beginning of the pulse, τ is the lifetime of the injected carriers and μ_p and b are the hole mobility and mobility ratio respectively. Thus, if $L = L_1 + L_2$, and consequently $R_1^2 L_1^{-2} = (R_1 + R_2)^2 L^{-2}$, then

$$\delta V = \frac{\{I_s(R_3 + R_4) + \delta I_e R_4\}(R_1 + R_2)^2 \gamma_e \mu_p (b+1) \tau \delta I_e \{1 - \exp -(t/\tau)\}}{(R_1 + R_2 + R_3 + R_4) L^2}. \qquad \ldots\ldots(11)$$

The value of τ can be found by observing the time dependence of δV and thus, as all the other quantities are either known or can be measured, the value of γ_e can be calculated.

2.2. *Emitter Current applied as Square Pulses superimposed on a Steady D.C. Component*

The circuit for the alternative procedure is the same as that used for the above method except that the battery E_2 is inserted. This battery gives rise to a constant emitter current I_e through the contact, and the bridge is initially balanced so that $R_1' R_4 = R_2 R_3$. This leads to

$$I_1'(R_1' + R_3) - I_2'(R_2 + R_4) = E_1 \qquad \ldots\ldots(12)$$

$$I_1' + I_2' = I_e. \qquad \ldots\ldots(13)$$

The square constant current pulse δI_e is then applied so that R_1' becomes R_1'' and consequently I_1' changes to I_1'' and I_2' changes to I_2''. Under these conditions

$$I_1''(R_1'' + R_3) - I_2''(R_2 + R_4) = E_1 \qquad \ldots\ldots(14)$$

and

$$I_1'' + I_2'' = I_e + \delta I_e. \qquad \ldots\ldots(15)$$

The voltage change across the ends of the filament is now given by

$$\delta V = R_3(I_1'' - I_1') - R_4(I_2'' - I_2'). \qquad \ldots\ldots(16)$$

This expression can be reduced to

$$\delta V = \frac{\{I_1'(R_3 + R_4) + \delta I_e R_4\}(R_1 + R_2)^2 \gamma_e \mu_p (b+1) \tau \delta I_e \{1 - \exp(-t/\tau)\}}{(R_1 + R_2 + R_3 + R_4) L^2} \qquad \ldots\ldots(17)$$

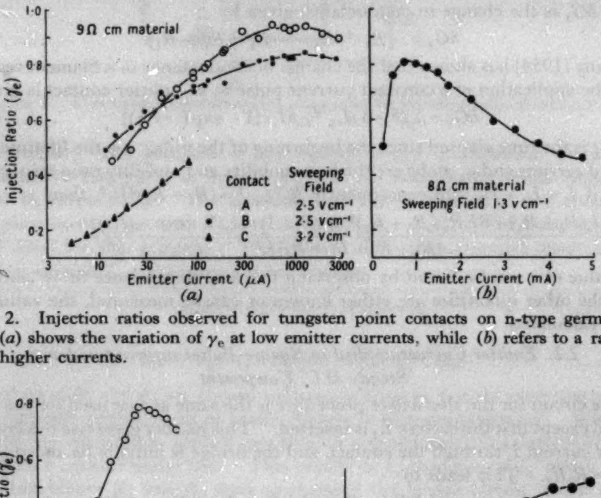
and thus γ_e can be evaluated.

§ 3. APPLICATION OF THE METHOD TO TUNGSTEN–GERMANIUM CONTACTS

Experiments were carried out on germanium surfaces which had been freshly etched with CP4 etch (Haynes and Shockley 1951). The specimens used were approximately 2 cm long and 1 mm × 1 mm in cross section. The tungsten contacts were made from 0·005 inch diameter wire which had in each case been pointed by dipping in boiling sodium nitrite. The measurements were made in a lightproof box and the ambient air was dried with phosphorus pentoxide.

3.1. *Results*

The value of γ_e over a range of emitter currents is shown for n-type and p-type germanium in figures 2 and 3 respectively. The observed variation of γ_e with emitter current at low currents is expected to give rise to a corresponding change in current gain if the arrangement is considered as a filamentary transistor.



(*a*) (*b*)

Figure 2. Injection ratios observed for tungsten point contacts on n-type germanium. (*a*) shows the variation of γ_e at low emitter currents, while (*b*) refers to a range of higher currents.

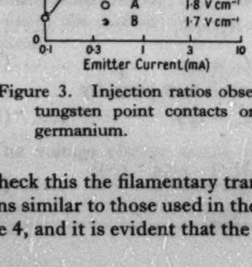


Figure 3. Injection ratios observed for tungsten point contacts on p-type germanium.

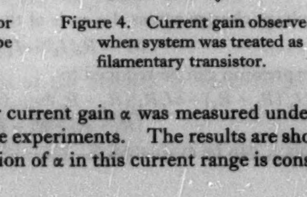


Figure 4. Current gain observed when system was treated as a filamentary transistor.

To check this the filamentary transistor current gain α was measured under conditions similar to those used in the above experiments. The results are shown in figure 4, and it is evident that the variation of α in this current range is consistent with the changes observed in γ_e.[†] Other published results of current gain (see, for example, Hogarth 1954) do not show this variation. This is ascribed to a difference in surface conditions. Surface conduction in a channel is probably increasingly important as higher purity germanium is used as in the present experiments.

As indicated above, in order to evaluate γ_e it is necessary to find τ from the time dependence of δV. It might be expected that, for a given specimen, τ would be a constant. However, a small change in τ was observed when the sweeping field was varied (emitter current constant). The value of τ was found to decrease with increasing field. This effect is ascribed to changes in the configuration of the electric field inside the specimen. At high sweeping fields, as a result of these changes, the injected carriers would be swept down the filament close to the surface and consequently recombination would be more rapid. An increase in emitter current was found to result in an increase in τ, for constant sweeping field, and this is probably due to a similar mechanism.

3.2. *Discussion*

The assumptions on which the method is based lead to opposing requirements concerning the magnitude of the sweeping field. A low sweeping field is required so as to prevent injected carriers being swept out of the filament through the end contact. On the other hand, a high field is necessary in order to minimize the relative importance of longitudinal diffusion effects and to prevent any large accumulation of injected carriers under the emitter contact. In the present system it is difficult to estimate the minimum field necessary to satisfy these conditions. In the present experiments, therefore, the sweeping field was kept as high as possible, subject to the condition that the fraction of the injected carriers swept out of the filament never exceeded 2%. Measurements carried out at constant emitter current showed that the observed value of γ_e increases with increasing sweeping field. This may have been due to an actual change in γ_e arising from an alteration in the conditions under the injecting contact. Alternatively, the change in the observed γ_e may have arisen because the above assumptions concerning longitudinal diffusion and accumulation of carriers under the contact were not completely satisfied. At lower sweeping fields this effect was important but under the conditions to which figures 2 and 3 refer the changes with field were very small.

Figures 2 and 3 show that γ_e increases with increasing emitter current for low currents, reaches a maximum at a current of the order of 1 mA, and then decreases with increasing emitter current. The latter part of this variation, i.e. the decrease of γ_e with increasing current, has been discussed by Banbury and Houghton (1954). The present results in this current range are consistent with the measurements reported by these authors. This variation is attributed to changes in the local hole concentration under the injecting contact. The mechanism of the increase of γ_e with current at low emitter currents is not yet understood. It may be associated with conduction processes in the inversion layer ('channel'). Those holes taking part in surface conduction which survive recombination may subsequently be injected into the bulk of the specimen. If this is the case, the present results would indicate that recombination is appreciable for very small emitter currents and that delayed injection becomes increasingly important as the emitter current is increased. γ_e would thus increase until the process discussed by Banbury and Houghton becomes significant. The variation of γ_e at low emitter currents is in the wrong direction to be accounted for by a breakdown of the above assumptions involving the magnitude of the sweeping field.

[†] The analysis of the filamentary transistor has been given by Shockley, Pearson and Haynes (1949). The quantity α_0 is the current gain of the resistor network which would be obtained if there were no transistor action.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for placing research facilities at his disposal, Drs. H. K. Henisch and P. C. Banbury for encouragement and valuable discussions, and the New Zealand Defence Research Organization for financial support.

REFERENCES

BANBURY, P. C. and HOUGHTON, J. 1954, *Physica*, **20,** 1050.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81,** 835.
HOGARTH, C. A., 1954, *Proc. Phys. Soc.* B, **67,** 636.
MANY, A., 1954, *Proc. Phys. Soc.* B, **67,** 9.
SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R. 1949, *Bell Syst. Tech. J.*, **28,** 344.

Measurements of the Sound from Circular Cylinders in an Air Stream

By J. H. GERRARD
Fluid Motion Laboratory, University of Manchester

Communicated by P. R. Owen; *MS. received* 19*th July* 1954 *and in amended form* 9*th March* 1955

Abstract. The frequency and intensity of the sound produced by flow past circular cylinders was investigated experimentally in the range of Reynolds number in which a periodic shedding of vortices occurs. The cylinders were placed in a wind tunnel or on a whirling device.

The sound resulting from the production of a vortex street was found to possess a dipole field which consisted of a predominant fundamental frequency accompanied by harmonics. The investigation was confined to the examination of the intensity and frequency of the fundamental. When this frequency was expressed as a Strouhal number Nd/U agreement was obtained between the acoustic frequency and that found, by other workers, in the wake itself. The sound intensity was found to vary with Reynolds number in a similar way to the pressure coefficient at the rear of the cylinder.

§ 1. Introduction

The acoustical phenomena accompanying the flow of a gas but not involving vibration of solid bodies or cavity resonance may be termed aerodynamic sound. Whether this sound is a pure tone, or has the continuous frequency spectrum characteristic of noise, depends upon the Reynolds number of the flow. Three ranges of Reynolds number may be defined for the flow of a gas past rigid bodies. At Reynolds numbers between 50 and 10^5 the periodic shedding of vortices takes place; the Kármán vortex street is the notable phenomenon of the lower part of this range. At Reynolds numbers below 50 no vortices are shed, and at Reynolds numbers greater than 10^5 the periodic characteristics of the flow are replaced by the random fluctuations of turbulence. Flow in the intermediate range of Reynolds number can give rise to pure tones, such as aeolian tones, jet tones, and edge tones. These may conveniently be termed 'aerodynamic tones'. Aeolian tones, the first to be investigated (Strouhal 1878), are the acoustical concomitants of the shedding of vortices behind a bluff obstacle in a fluid stream.

Sound of an essentially pure tone character is radiated from the flow of air past a circular cylinder in the range of Reynolds number from 50 up to at least 3×10^4. The Kármán vortex street is only formed at the lower Reynolds numbers of this range. It follows that the pressure fluctuations responsible for the sound are situated close to the cylinder and are associated with the production and breaking away of vortices. In the theory of the aerodynamic production of sound Lighthill (1952) shows that the fluctuating forces of an air flow on a rigid body may be expected to possess an associated dipole radiation field. The experiments described in this paper have shown that this interpretation of the production of sound is correct. It is demonstrated that the sound is not produced by cylinder vibration. The radiation is found to possess a dipole field of an essentially pure tone character; the harmonics present are considerably less intense than the fundamental frequency. Measurements were made of the sound produced when the cylinders were placed in a wind tunnel, and when they were moved through the air on a whirling device similar to that used by Strouhal and later workers (Relf 1921).

Much effort has been devoted, in the past, to the determination of the frequency of aeolian tones and the frequency of vortex shedding. Relf (1921) showed that the acoustic frequency was the same as that of vortex shedding (from one side of the cylinder) and then only in the range of Reynolds number from 100 to 500. It has been usual to express the frequency in terms of Strouhal number (Nd/U) and to investigate the variation of this with Reynolds number. The work prior to 1937 is summarized graphically in this manner by Lehnert (1937). The thirteen different curves of various experimenters show a considerable amount of scatter. The three most recent determinations of vortex-shedding or wake frequency by Lehnert (1937), Kovasznay (1949), and Roshko (1953) are in very good agreement. The author's acoustic determinations are in agreement with these in the range of Reynolds number from 300 to 10,000.

The sound intensity I measured in the plane perpendicularly bisecting the cylinder axis (the ' central plane ') was found to obey the relation

$$I = \frac{\overline{p^2}}{\rho_0 a_0} = \frac{\rho_0 U^4}{a_0} \frac{l^2}{r^2} \cos^2\theta f(R), \qquad \ldots\ldots(1)$$

where $\overline{p}^2$ is the mean square sound pressure, ρ_0 the density of air, a_0 the velocity of sound in air, U the velocity of the cylinder relative to air, l the cylinder length, r the distance from the midpoint of the cylinder to the point of observation, and $R = Ud/\nu$ Reynolds number. ν is the kinematic viscosity of air.

The variation of intensity with Reynolds number is very similar to that of the pressure coefficient measured at the rear of a cylinder. Equation (1) is not what is produced by simple theory. The reason for this is thought to be the complicating effect of a spanwise periodicity which is found in the vortex shedding. The intensity of the sound radiated from whirled cylinders was investigated by Holle (1938); his results are compared with those of the author.

§ 2. APPARATUS AND PROCEDURE

Measurements have been made with the cylinders in a low turbulence wind tunnel (Collis 1952) and on a ' whirler ' which moved them on a circular path perpendicular to their axes. The whirler axle and the radial arms supporting the cylinders were made of $\frac{3}{8}$ in. diameter steel rod. The driving motor had a constant speed of 25 revolutions per second. The velocity of the cylinder in the whirler was varied by altering the distance of the cylinders from the axle. Two cylinders were used, one at each end of the radial arms. The microphone was placed in the central plane. In the wind tunnel measurements, twenty cylinders were spaced 2·5 cm apart in a plane across the 50 cm square working section of the tunnel. The cylinders were 30 cm long and were supported between $\frac{1}{2}$ in. diameter cylinders which spanned the tunnel. The microphone was placed by the side of this grid and about 3 cm downstream of the line perpendicularly bisecting the cylinders in the plane of the grid. At this section the tunnel is open to the atmosphere for a distance of about 6 cm.

The wind tunnel was calibrated by measuring the air velocity in the position of one of the cylinders, the other 19 remaining in position. In the whirler measurements uncertainty arises from the unknown effect of the cylinder moving into the wake of the preceding cylinder. The distance the wake moves outwards depends upon draughts and is therefore not constant. It was assumed, therefore, that the absolute speed of the cylinder was that due to rotation at 25 cycles per second (correct to 2 or 3%) and that the air was stationary. Comparison with the Strouhal number Nd/U determined from the wind tunnel measurements served to correct the velocity to the true value. This method has the effect of transferring errors in frequency, the magnitude of which can be seen from the scatter of the points in figure 3, to errors in velocity.

The microphone used was an electromagnetic one (S.T.C. 4021 E). The signal from this was amplified, passed through one of twenty-seven $\frac{1}{3}$-octave filters, and after further amplification the level was measured with a square law device (vacuo-junction and galvanometer). In all cases, after the signal had been measured with the cylinders in place, the measurement was repeated with all conditions identical except that the cylinders were removed. The sound radiated by the cylinders was then calculated and plotted as a function of frequency (see figure 2). The 27 filters possess charcteristics of nearly identical shape; the curve upon which the observations are expected to lie if the signal is a pure tone of frequency N is therefore the mirror image of the filter characteristics and has its maximum at this frequency. The fundamental frequency and the signal intensity were determined by fitting a curve of this shape to the observations.

It will be shown that the sound field in the central plane of the cylinder can be represented (at large distances) by $(A\cos^2\theta)/r^2$, where θ is the angle which r makes to the direction of U. In the whirler experiments θ and r are functions of time, and it may be shown that

$$\overline{\left(\frac{A\cos^2\theta}{r^2}\right)} = \overline{A\cos^2\theta}\,.\,\overline{\left(\frac{1}{r^2}\right)}, \qquad \ldots\ldots(2)$$

where the bars denote means with respect to time. The left-hand side of equation (2) is measured in the experiments and thus the effective distance is the harmonic mean of r^2. Intensities measured in the wind tunnel experiments are not included in this paper because of the difficulty of reduction of results mainly due to the sources of sound being enclosed in the wind tunnel.

§ 3. DIRECTIONAL PATTERN OF THE SOUND FIELD

By plotting the directional distribution of the intensity of the sound emitted in the central plane from the flow round a cylinder it was found that the field was predominantly dipole in character. The directional characteristics were determined from photographs of the cathode-ray oscillograms of the microphone signal when only one cylinder was revolved on the whirler. One directional pattern is shown in figure 1, where it is compared with curves proportional to $\cos^2\theta$. One lobe of the dipole is divided into two because of the shadow cast by the axle of the whirler.

The intensity possesses large and irregular fluctuations of longer time scale than the period of vortex shedding. This is in evidence in figure 1 where the two lobes on the same side are of widely differing magnitude. These fluctuations were also present in the wind tunnel measurements. The signal from a hot wire anemometer placed in the wake close behind the cylinder also showed similar fluctuations except at the lowest Reynolds numbers.

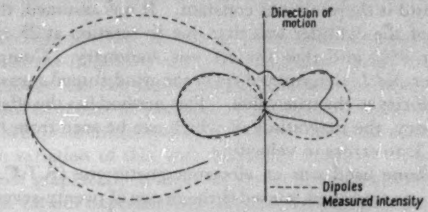


Figure 1. Directional pattern of the sound field.

§ 4. FREQUENCY DETERMINATION

It is important to explore the whole frequency range for it is not expected that the sound produced by the wake of a circular cylinder will be at a single frequency. The sound was found to consist of a predominant fundamental and some harmonics. In all cases the harmonics were considerably less intense than the fundamental, so much so that the definite existence of only the third harmonic can be deduced from the measurements.

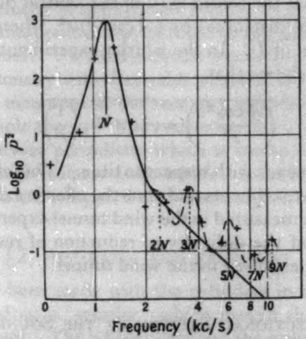


Figure 2. Sound spectrum.

Figure 2 shows one set of experimental results. The full line indicates where experimental points would be expected to lie if the signal was a pure tone of the intensity and frequency indicated. This curve is adjusted to give the best fit for points near the fundamental frequency. The presence of harmonics is indicated by the fact that observations lie above the curve at the high frequency end. Many sets of harmonics could obviously be fitted to these points. One way in which this might be done is shown in the figures by dashed lines. The harmonic content has not been investigated in detail. The remainder of the work concerns only the fundamental frequency and, in the whirler experiment, the intensity of this sound.

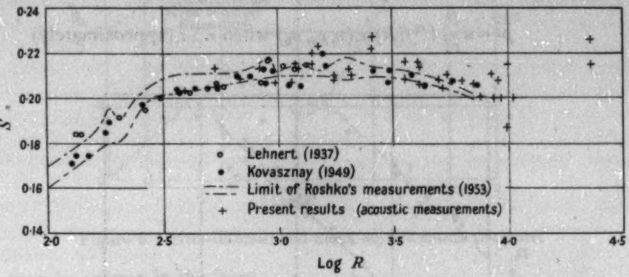


Figure 3. Non-dimensional sound frequency compared with wake frequency.

The dependence of frequency upon the various parameters is best portrayed by computing the Strouhal number $S=Nd/U$ and plotting this as a function of Reynolds number. This is done in figure 3 where the Strouhal numbers determined from the wind tunnel measurements are compared with the latest determinations of the Strouhal number based on frequency determined within the wake. It is seen that good agreement is obtained, thus extending the confirmation by Relf (1921) at low Reynolds numbers of the equality of the acoustic and wake frequencies.

§ 5. INTENSITY DETERMINATIONS

The experimental determination of the law relating the mean square sound pressure $\overline{p^2}$ to the various parameters of the problem is best approached from a dimensional consideration. The sound pressure can be expressed as a function of all the variables upon which it may be expected to depend, thus

$$\overline{p^2}=g(n, l, r, d, U, \nu, \rho_0, a_0),$$

where n is the number of cylinders used, d is the cylinder diameter; in these experiments, the last three parameters are essentially constant. The intensity is independent of the elastic constants of the cylinder since it was found that cylinder vibration played no essential part in the production of sound; thus experiments with the small diameter cylinders slack and taut produced no significant difference in intensity. Also, intensity from a vibrating cylinder depends upon the square of cylinder density (Morse 1948), but when experiments were performed in which only the cylinder density was altered the measured intensity remained unaltered. Cylinders of steel (density 7·8 g cm^{-3}) and aluminium (density 2·8 g cm^{-3}) as well as ones of hollow brass were used.

The distance r from the cylinder to the point of observation undergoes a cyclic variation and its effective value depends upon the way in which $\overline{p^2}$ varies with r. To start with, therefore, it was assumed that $\overline{p^2}\propto 1/r^2$. Experiment showed (see figure 4) that $\overline{p^2}\propto l^2/r^2$ for the observations made at the larger values of the effective distance $\bar{r}$. Retaining only those observations for which $\overline{p^2} \propto l^2/\bar{r}^2$ it was found that the mean square sound pressure was proportional to U^4 (figure 5) and to a function of Reynolds number which has the form of the curve in figure 6. It was assumed that $\overline{p^2}$ was proportional to the number of cylinders ($n=2$) used in the experiments. We may thus write

$$\overline{p^2} = n\frac{l^2}{\bar{r}^2}U^4 f(R) g, (\nu, \rho_0, a_0) \text{ when } \bar{r} > l \text{ (approximately)}$$

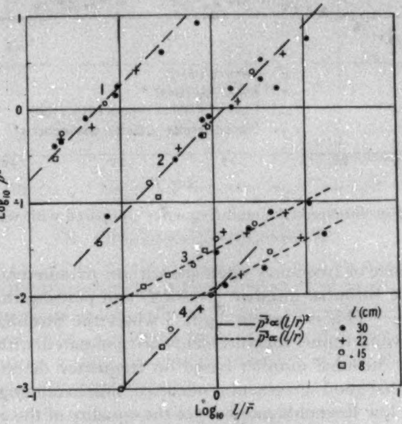


Figure 4. Sound intensity as a function of l/r.

Curve	Log R	Log U (cm sec^{-1})	d (cm)
1	3·80	3·46	0·317
2	3·57	3·21	0·318
	3·44	3·30	0·204 (l=30 cm)
3	3·15	3·23	0·121
4	2·56	3·06	0·049

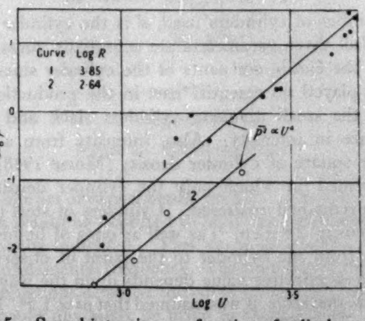


Figure 5. Sound intensity as a function of cylinder velocity.

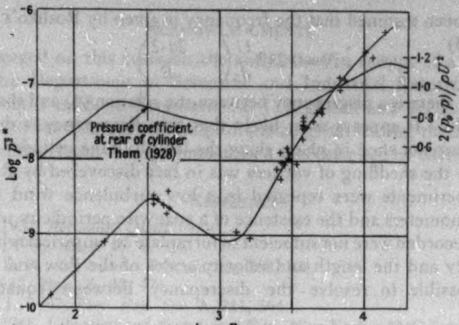


Figure 6. Non-dimensional mean-square sound pressure.

which may be written in the form

$$\frac{\overline{p^2}}{n\rho_0{}^2U^4}\frac{\bar{r}^2}{l^2}=f(R)g_2(\nu,\rho_0,a_0),$$

and as the left-hand side is non-dimensional g_2 must be non-dimensional, and must therefore be a numeric as no non-dimensional combination of ν, ρ_0 and a_0 exists. Any measurements (in the range of $\bar{r}$ where $\overline{p^2}\propto l^2/\bar{r}^2$) may therefore be reduced to a universal function of Reynolds number by constructing the expression

$$\overline{p^2}^*=\frac{\overline{p^2}\,\bar{r}^2}{n\rho_0{}^2U^4l^2}=f(R), \qquad \ldots\ldots(3)$$

which is plotted (with $n=2$) in figure 6.

Measurements close to the whirler showed that $\overline{p^2}\propto l/\bar{r}$ where $\bar{r}$ is the effective distance determined from the harmonic mean of $1/r$.† The value of $\overline{p^2}^*$ determined from these measurements agreed well with the curve in figure 6, which was determined only from equation (3).

§ 6. Discussion of Results

It has been found that at large distances from a cylinder in a flow of velocity U relative to the cylinder the sound intensity is given by

$$I=\frac{\rho_0U^4}{a_0}\frac{l^2}{\bar{r}^2}\cos^2\theta f(R). \qquad \ldots\ldots(1)$$

The pendulation of the wake behind the cylinder is accompanied by a periodic force on the cylinder which from Lighthill's theory (1952) is expected to produce a dipole field. The intensity produced in the central plane can be calculated from equation (12) of Lighthill's paper. The fluctuating lift force on the cylinder is the only one of the same frequency as the sound. By considering the cylinder as a line dipole and assuming that the fluctuating lift is in phase along the length of the cylinder we obtain for the intensity in the central plane,

$$I=\frac{\rho_0U^6}{a_0{}^3}\frac{l^2}{\bar{r}^2}\cos^2\theta f(R), \qquad \ldots\ldots(4)$$

† This effective distance was not significantly different from that determined from the harmonic mean of $1/r^2$.

where it has been assumed that the frequency is given by Roshko's experimental relation (1953)

$$N = 0{\cdot}212\frac{U}{d}\left(1 - \frac{21{\cdot}2}{R}\right).$$

It is seen that there is a discrepancy between the relation (4) and that determined experimentally. It appears most likely that this discrepancy is due to the fact that vortices are not shed in phase along the length of the cylinder. A spanwise periodicity in the shedding of vortices was in fact discovered by Roshko (1953). Roshko's experiments were repeated in a low turbulence wind tunnel, using hot wire anemometers and the existence of a spanwise periodicity was confirmed, but the data recorded were not sufficient to formulate an empirical relation between the periodicity and the length and velocity scales of the flow and it was, therefore, not possible to resolve the discrepancy between equation (1) and equation (4).

Holle (1938) investigated the dependence of sound intensity upon the velocity for various values of cylinder diameter. He found no systematic dependence upon diameter except at high velocities ($U > 30\,\mathrm{m\,sec^{-1}}$) where the intensity was found to be roughly proportional to d. For these high velocities is given the empirical relation

$$I = 5 \times 10^{-10}\frac{U^7 dl}{D^2}\ \mathrm{ergs\,sec^{-1}\,cm^{-2}},$$

where D is the distance from the point of observation to the whirler axle, and the parameters in the equation are in c.g.s. units. A graph showing the relation between I and l is well represented by $I \propto l^2$. When Holle's results are reduced to the non-dimensional sound pressure $\overline{p^2}^*$ (assuming that equation (1) is obeyed) the results agree with the curve of figure 6.

The way in which the sound intensity varies with Reynolds number is correlated with the experimental results of other workers. In figure 6 the pressure coefficient computed by Thom (1928) from the pressure measured at the rear of a cylinder is compared with the variation in non-dimensional sound intensity. The pressure coefficient is seen to possess a similar variation with Reynolds number as does the sound intensity. Kovasznay (1949) and Roshko (1953) both found that at Reynolds numbers below 150 the vortex street was stable and regular and no turbulence was present in any of the vortices. At higher Reynolds numbers instability and irregularities occurred in the street. Roshko defines an 'irregular range' of vortices for Reynolds numbers between 150 and 300. In this range bursts of turbulence were observed by Roshko at a point six diameters behind the cylinder. It appears that between these Reynolds numbers a transition occurred: the vortices changed from all laminar to all turbulent. It is possible that there is a connection between Roshko's irregular range and the range of Reynolds numbers between about 300 and 1000 where some sort of transition appears to occur in the sound intensity. Roshko expresses the opinion that changes in main stream turbulence may be expected to change the position of the irregular range.

The frequency of vortex shedding from one side of the cylinder and the acoustic frequency have been found to be equal. The results given in the paper extend the range of Reynolds number over which Relf found this to be true. Fluctuations in the drag force which have the frequency of the second harmonic are known to be smaller than the fluctuations in lift and would therefore be expected to produce a dipole field of smaller intensity.

ACKNOWLEDGMENTS

Whilst engaged on this research the author was in receipt of a maintenance grant from the Department of Scientific and Industrial Research. Grateful acknowledgment is made to Professors M. J. Lighthill and W. A. Mair and to Mr. P. R. Owen for interest and helpful advice given during the course of the work. Most of the electronic equipment used was designed by Mr. V. H. Attree.

REFERENCES

COLLIS, D. C., 1952, *Aeronaut. Quart.*, **4**, 93.
HOLLE, VON W., 1938, *Akust. Z.*, **3**, 321.
KOVASZNAY, L. S. G., 1949, *Proc. Roy. Soc.* A, **198**, 174.
LEHNERT, R., 1937, *Phys. Z.*, **38**, 476.
LIGHTHILL, M. J., 1952, *Proc. Roy. Soc.* A, **211**, 564.
MORSE, P. M., 1948, *Vibration and Sound*, 2nd Edn (New York : McGraw-Hill), p. 299.
RELF, E. F., 1921, *Phil. Mag.*, **42**, 173.
ROSHKO, A., 1953, *National Advisory Committee on Aeronautics*, Tech. Note 2913.
STROUHAL, V., 1878, *Wied. Ann.*, **5**, 216.
THOM, A., 1928, *Aeronautical Research Council, Reports and Memoranda* No. 1194.

A Study of Powder and Granular Ridges in a Sound Field

By MARY D. WALLER
Formerly at the Physics Department, Royal Free Hospital of Medicine

MS. received 20th December 1954 *and in amended form 9th February* 1955

Abstract. The author's vibrating plate method of forming powder ridges on a flat surface has now been used to study the phenomenon with graded powders of different densities and at different frequencies and intensities of vibration of the sound field. It is found that the distances between the ridges increase roughly as the square root of the particle diameters, a result which is in agreement with Bagnold's observations with powders under oscillating water, but needs still to be taken into account of in connection with Kundt powder ridges. 'Granular' ridges, one particle wide, are nearer together than powder ridges, the spacings of which increase with the amount of powder which is present. Much larger intensities are required to move dense powders than light powders; the resulting spacings are, however, approximately the same. The variables which have most influence on the spacing are the diameters of the grains and the frequency of the vibration, not the intensity. These matters are discussed in relation to the phenomenon of 'wavy' ridges. Experiments in which larger objects, as well as powders, are introduced into the field are also described.

§ 1. INTRODUCTION

THESE experimental studies were undertaken with the purpose of arriving at some conclusions regarding the many variables which affect the formation of powder ridges in a sound field.

In the experiments a new technique described by the author (Waller 1954 b) has been used. A circular plate is supported a few millimetres above a horizontal surface on which the powder is scattered uniformly. When the plate is set into vibration, by striking it with a suitable hammer, in the mode in which a single nodal circle is formed, the air in the gap will vibrate radially; when the centre of the plate is moving downwards and the periphery upwards, air will be forced outwards and during the other half-cycle, air will move inwards. Thus the frequency of the air vibration will be equal to that of the plate. At suitable intensities circular powder ridges will be formed as shown, for example, in figure 7, where the three studs which support the plate at the nodal circle are also visible. The mean distance between ridges is estimated from measurements made on the main surface of the plate (not too near the centre or the periphery). As in Kundt's tube, too great an intensity causes large movements and accumulations of powder and such intensities have been avoided. It has been assumed that the radial field in the narrow air-gap is dominant, and that any effects of general circulations (Faraday 1831, Rayleigh 1884) are negligible.

It was soon realized that the quantity of powder used, as well as its grade, influenced the spacing of the ridges and that these might be divided into three types: (i) granular ridges which are one particle wide, obtained only when the powder is used sparingly (e.g. figure 9)†, (ii) powder ridges (e.g. figure 7), which become thicker and further apart as the amount of powder present increases, (iii) so-called ' ripple marks ', formed on whole beds of the powder.

The present studies are restricted to granular and powder ridges; ripple marks, which are also affected by the thickness of the powder bed have not, as yet, been investigated in any detail.

Andrade (1931, 1932) has shown that the occurrence of Kundt powder ridges depends upon the presence of aerial vortices. In his theory these leave a ridge on either side, parallel to the vibration vector, returning from above; the distance between two ridges is determined by the two vortices just meeting.

But even about one *fixed* object in an oscillating field, the sizes, shapes, strengths and senses of the circulating currents are variable and sometimes complex, cf. Carrière (1929), Andrade (1931, 1932), Skogen (1952), Andres and Ingard (1953). While ridges are being formed, numerous particles are set into movement, and since gravitational forces are now also operative, the situation is one of great complexity. Thus Darwin (1883) illustrated complicated systems of vortices in his paper on the formation of sand ripple-marks under oscillating water. Again, Bagnold (1946), working on a similar subject, has made a distinction between rolling-grain and vortex formative mechanisms. Bagnold has also reached conclusions, to be referred to later, concerning the important part which is played by the size of the particles.

The present writer (using glass plates which are not as good vibrators as brass plates and show troublesome electrostatic effects with some powders) has observed that while light particles leave the surface and appear to rotate in the senses of Andrade's vortices, denser particles do not leave the surface and appear to move to their equilibrium positions in a contrary direction. The ridges form almost instantaneously and this matter needs to be studied by stroboscopic methods.

In view of the above facts it has seemed worth while to collect and grade many different powders and to examine the behaviour of grains of various sizes, shapes and densities in the sound field, this being a subject which has been strangely neglected in studies with Kundt's tube. The possibility of sprinkling the powders either sparingly or in profusion upon the flat surface, and the further circumstance that adjustments to conditions of resonancy are not required, has made the carrying out of a rather arduous programme practicable.

§ 2. POWDERS, GRAINS AND IMMOVABLE OBJECTS IN A SOUND FIELD

2.1. *Grading and Classification of the Powders*

All heterogeneous powders were graded by passing them through a large number of sieves particulars of which are given at the top of table 1.

The size of grain, third row, has been specified as a ' diameter ' which, as is very usual, has been defined in terms of the mean size of the two apertures which isolated the grade. Since, as illustrated in figure 11, elongated particles align themselves parallel to the ridge, this definition of size is more accurate than might at first sight be supposed. For it is the height and breadth, rather than the length of the particles, which determines the distance between the ridges.

Powders, e.g. sand, the particles of which are hard and smooth and generally rounded, may be classified in terms of density and grain diameter. But when the particles are porous, of irregular shape and rough surfaced, a knowledge of the *bulk specific gravity* gives more information concerning the behaviour of a powder in moving air than the density of the solid matter. This was determined by weighing a given volume of the powder which was shaken down, but not pressed down, in a flat bottomed test-tube. Great accuracy is not required and since the shape of the particle is usually a characteristic of a given material it does not matter if the determination is made with a rather coarser or finer grade than the 40–60 grade which was chosen when available. The values, together with those of the corresponding specific gravities, are shown in table 1. To instance cork powder, the bulk specific gravity is only 0·07 as compared with the specific gravity of approximately 1*, for the powder, like all the seeds listed in table 2, just sank when shaken up in water, but floated in brine.

---

† For figures 1–17 see Plates I and II and for index to figures see table 3.

Table 1. Selected graded Powders, air-gaps 5 mm, 867 c/s

Mesh (holes per in.)	20	30	40	60	80	100	120	160	200	240	300
Aperture (μ)	915	592	414	261	191	114	100	80	73	65	54

Mean diameter (μ)			750	500	340	220	170	130	100	80	70	60
Powder	Bulk sp. gr.	Sp. gr.	Approximate average distance between ridges (mm)									
1	0·07	∼1	4–6	4–5	3–4	3	2·5 3·8	1·5 3	w			
2	0·25	∼1	4	4–3	2·5	2·3	1·7 3	1·3	w			
3	0·26	∼1	4	4–3	3	2·5	1·7 3	—	—			
4	0·5	∼1	3·4	3·2	2·6	2·2	2	1·5 2·2	w			
5	0·6	1·4	3·5	3	2·7	2	1·8	1·6 2·5	w			
6	0·8	*1·3*	3·3	3	3	2·4	1·4 3	2·7	1·1	1·1	w	w
7	1·3	2·2	—	2·7	2·2	2·1	2·1	2	1·8	1·3	w	
8	1·4	2·6	3·2	3	2·7	2·2	1·8	1·6	1·4	1·4	1·2	w
9	1·7	2·6	3·3	3	2·7	2	—	—	—	—	—	—
10	2·1	8·5	3·7	3	2·5	2·2	1·8	1·5	1·1	w		
11	5·0	11	—	2·7	2·7	2·3	2	1·6	1·8 2·7	w		

1 Cork, 2 Sawdust, 3 Tobacco ash, 4 Wood charcoal, 5 Coal, 6 Plastic spheres (diakon), 7 Table salt, 8 Sand, 9 Glass spheres, 10 Brass, 11 Lead.

w, wavy ridges (§ 2.7).

2.2. *Effects of Size of Particle and Quantity of Powder on the Spacing. Bulk Specific Gravity and Intensity*

The outstanding feature of experiments with graded powders is that, as illustrated in figure 1, for 14–18 grade sugar, coarse powders form widely separated ridges, while with finer material, see the 120–160 sand of figure 2, the ridges are much nearer together. Changes of density affect the spacing very little, as can be appreciated from the columns of table 1. It is, however, noteworthy that much larger intensities are needed to form ridges with dense material than with light material such as cork. The vertical movements of light materials are also much greater; thus cork ridges have been recognized in air-gaps 50 mm wide, but sand ridges have not been detected beyond 10 mm.

---

* The specific gravity of 'cork' is given as ∼0·25 in tables of constants and this value is sometimes to be found in theoretical work on moving powders.

Further data concerning the relation between the distances between ridges and the mean diameters of the grains are given in the two tables. The most reliable measurements are those derived from granular, or at any rate thin, ridges. The first three powders of table 1 are so light that they are easily moved into thicker, and therefore more widely, spaced ridges. This explains why the numbers in the early rows of the table are less constant then in the remaining rows.

Figure 18 was constructed from the results given in table 1 and table 2 (last column). It shows that over a large range of diameters, the ridge spacing varies *roughly* as the square root of the particle diameter. This conclusion is the same as Bagnold's (1946), and it is evident, therefore, that grain diameter is of fundamental importance in all powder ridge phenomena. Bagnold also concluded that as regards the repetition distances of his powder ripples under oscillating water, changes of density had but little influence (see also the remarks by Waller (1954 a) concerning Kundt's (1866 a) powder *Rippen*).

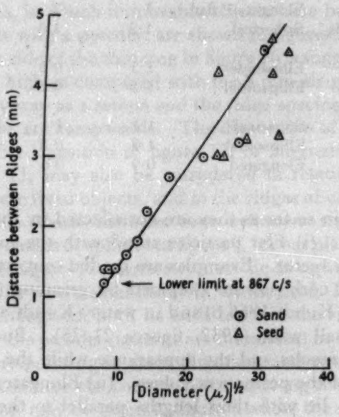


Figure 18.

2.3. *Ratio of Ridge Spacing to Particle Diameter*

Since the separation between adjacent ridges varies approximately as the square root of the particle diameter, it follows that the ratio of the separation to the particle diameter increases as the particle diameter decreases. Thus in figure 9 (× 3) the ratio is approximately 3 as compared with approximately 8 for the smaller particles of figure 10 (× 16). (A table of ratios may be easily constructed from table 1.) The vortices, which we may imagine present, presumably assume different shapes and relative sizes according to the diameters of the particles which initiate them, and the present results are therefore not without interest in connection with the many observations which are now being made of vortices around fixed objects in oscillating fields. References to some of these were given in § 1 and it may be noted that the senses of Andrade's vortices are reversed as compared with those of Carrière, that Andres and Ingard illustrate very different vortices at high and low values of Reynolds number respectively, and that in Skogen's figures inner vortices have the same senses as those of Carrière, while the senses of the outer vortices are those of Andrade (cf. also the senses of the vortices on one side of a fixed obstacle, at zero frequency, e.g. Prandtl (1952 chap. 3)).

2.4. *Alignment of Different Shaped Particles*

(*a*) Shape of particle, as is evident from tables 1 and 2 in which spheres, cubes cylinders, ellipsoids and generally elongated grains and particles of irregular shape are represented, does not have any noticeable effect upon the distances between the ridges.

Table 2. Larger Grains of Various Shapes, air-gap 5 mm, 867 c/s

Grain	Approximate shape	Length (mm)	Diameter (mm)	Grade	Mean distance between ridges (mm)
Spaghetti	Cylindrical	4	2		12
Turnip seed	Spheroidal	1·4	1·3	18–20	4·7
Lettuce seed	Elongated, flat	4	2	18–20	4·5
' Hundreds and thousands ' (Figure 9)	Spherical	1·3	1·3	18–20	4·2
Celery seed	Ellipsoidal	1·3	0·6	20–20	4·2
Virginia Stock seed (Figure 11)	Ellipsoidal	1·5	0·6	20–30	3
Sago cereal	Spherical	1·1	1·1	20–30	3·3
Timothy grass seed	Ellipsoidal	1·8	0·8	30–40	3·2
Poppy seed	Elongated	0·9	0·6	30–40	3

(*b*) The particles, in so far as they are not affected by their neighbours align themselves as follows: (i) Flat particles stand with the main surface at right angles to the vibration vector. Examples are graded soot, sawdust, tobacco ash, kitchen salt, pith and cork powder prepared by grinding (cf. Rayleigh's disc, vertical *Rippen* in air (Kundt 1866 b) and in water (Kundt and Leymann 1874), and Andrade's pith-ball walls (1932, figures 21–25). Buoyancy and gravity evidently affect these results and the appearance while the tube is sounding is not the same as that of the permanent ridges. (ii) Elongated particles, as noted in § 2.1 for figure 11, lie with their lengths parallel to the ridges. Examples are wood-charcoal powder and the cylinders and seeds listed in table 2. (iii) Cubes also set themselves so that the moment of inertia is least parallel to the ridge, i.e. diagonally as in figure 10 which shows fairly perfect cubes of table salt. It will be noted that each grain is separate, a result which may be contrasted with the closely touching spheres of figure 9.

(*c*) The well-known hydrodynamical ' attractions ' and ' repulsions ' between individual particles in a sound field may be conveniently studied with some of the elongated grains mentioned in table 2. For instance (i) A Timothy grass seed lying parallel to the vibration vector on the circle passing through a fixed stud (of figure 7) turned through a right angle so that the more pointed end touched the stud. (ii) One seed lying on the radius through the stud and touching it was usually repelled from it and assumed a transverse position. (iii) Two seeds touching in the radial position (not near a stud) assumed transverse directions end to end. This may be regarded as one ridge two particles long. (iv) Two cylinders placed radially end to end rotated and moved away from each other to form the equivalent of two ridges, one particle long. (v) When two cylinders were placed transversely and touching they were usually ' repelled ' to the same final positions as in (iv); an alternative result (with seeds) was that the particles remained unmoved, so forming one ridge two particles wide and one particle long. These few observations might well be followed up using, for example, cylinders cut from different gauge wires of various materials, placed on a polished surface to reduce the friction. Incidentally, cylinders do not roll about so easily as spheres and are more convenient for experiment. It is important (see § 2.3 above) to experiment with different diameters and it may be noted that in Cook's (1902, 1903) paper the conclusions reached regarding the forces were made with spheres of diameter 0·2 mm while those of Andrade (1932, § 5) were made with cylinders of diameter 1·5 mm.

2.5. *Immovable Objects*

It will have been noticed that the studs (see figure 7) affect the powder ridges in a way which reminds one of the behaviour of iron filings near a magnet. A few further observations, in which immovable objects have been introduced into the sound field together with a powder, are shown in figures 12 to 17. The absence of powder on either side of the thin ring in figure 16 is suggestive of large vortices; the ring was 7 mm high as compared with the 8 mm air-gap. The ring does not appear to act in any way as a screen and the ridge spacings (dried bread crumbs) outside it and inside, are comparable. The disposition of the sand around a cube placed in the ' stable ' position of figure 12, or alternatively in the ' unstable ' position of figure 13, may also be considered in relation to the formation of vortices around these fixed objects, and to the ridges of cubic grains in figure 10. Similarly figure 15, which includes four cylinders standing on end, may be compared with the chains of spheres in figure 9. A bar placed radially (not shown) has little effect; cf. figure 17 where two bars lie at right angles to the vibration vector, and figure 14 where a bar lies at 45° to the field. Later observations made with a bar, short enough for both its ends to be in the same field, show that, in this latter position, the powder patterns are the same at both ends; ridges converge towards the corners pointing in their direction, and powder is removed at two corresponding places near the corners pointing along the radius vector.

Special attention may be given to figure 8 where the quasi-movable cylinders are playing the double role of forming widely spaced ridges which can be seen to converge towards the fixed stud at the top, and of acting as ' obstacles ' to the much more nearly spaced sand ridges, which in their turn can be observed to converge on the ends of the cylinders.

2.6. *Minor Ridges*

The occurrence of two numbers in one column of table 1 indicates that, as in figure 5, minor ridges were present between two major ridges. Such ridges occur sometimes in Kundt's tube, and Andrade (1932, § 10·7 and figure 2) has concluded that " when the particles are not of uniform size, intermediate ridges tend to be formed by the lighter particles, which are swept by the air vortices to a position between the main ridges. With uniform spherical particles made of wax, no intermediate ridges can be formed ". The following observations, however, suggest, that the smaller particles constituting the minor ridges form their own independent vortices. (i) When a minor ridge is present the spacing appropriate to a given grade of powder is increased (cf. the sequence of numbers in the earlier rows of table 1 for the 220 μ, 170 μ and 130 μ diameter particles). (ii) When sufficient powder is present, even uniform spheres may produce minor ridges as in the case of the 170 μ diakon spheres mentioned in table 1 (these, being transparent, did not yield clear photographs). (iii) A powder of high density, which it is believed does not leave the surface, sometimes forms minor ridges (see the 100 μ lead powder, table 1). (iv) An artificial mixture of two widely separated grades of powder may form major and minor ridges as illustrated in figure 6. (v) More than one minor ridge can exist between major ridges. This was observed, for instance, with cork powder at 120 c/s. Figure 8 provides an extreme example of this possibility.

For all these reasons it has been concluded, that if vortices are responsible for main ridges, vortices, of appropriately smaller size, are also responsible for minor ridges.

2.7. *Lower Limit of Diameter at a Given Frequency. Waviness*

With the finest powders included in table 1 it became impossible at the given frequency, to obtain straight closely spaced ridges like those shown for 100 μ sand in figure 2. The 70 μ sand gave less definite results, see figure 3, while the 54 μ sand ridges (not shown) were wavy and *further* apart than in the preceding grade. With the still smaller 20 μ lycopodium spores the ridges were sinusoidal in character and widely spaced. Those obtained at a rather higher frequency of 1330 c/s are illustrated in figure 4. This new phenomenon, denoted by w in the table, appears not to depend upon density, and it disappears in favour of straight closely spaced ridges, provided the frequency is sufficiently raised. Thus, the mean ‘ crest to crest ’ distances in a ridge, at 617, 867 and 1330 c/s, were of the order 8, 5 and 4 mm respectively, the corresponding spacings being roughly 5, 4 and 2·5 mm. The present technique is not well adapted for high frequencies but it was evident that at 2900 c/s the lycopodium striations were already nearly straight, thin and very close together.

These results are important since they show that at a given frequency there is a lower limit of diameter below which the usual ridge phenomenon does not occur. Diameter and frequency must, to some extent ‘ fit ’ and the occurrence of very finely spaced ridges is dependent on both small diameter and high frequency.

§ 3. EXTERNAL CONDITIONS

So far, we have been interested in keeping the external conditions as constant as possible while studying the behaviour of different powders and grains. It now remains to describe experiments in which the intensity and frequency of vibration were altered while the powder remained the same. This subject has received special attention in connection with Kundt ridges for which, according to Andrade (1932, § 4), “ the amplitude has been cited as the criterion, since it is open to direct observation, but general considerations of the hydrodynamics of viscous motion would suggest that it is the velocity of the air past the particles, possibly the acceleration, that constitutes the criterion ”. Since the velocity varies directly as the amplitude and the frequency, it follows that an increase in either intensity or frequency should increase the spacing.

3.1. *Changes of Intensity. Comparison with Kundt Ridges*

All observers are agreed that within the restricted range of intensities which produce good ridges, the spacing increases with the intensity. But in the writer's experience the ratio of maximum to minimum spacing never exceeds two or possibly three. This conclusion was confirmed by observations in which the intensity was increased by using narrower air-gaps. Having studied all the published figures on the subject, especially those of Kundt (1866), Irons (1929) and Andrade (1932), the writer suggests that this is also true of Kundt powder ridges and that formulae which give the variation in spacing between antinode and node do not accord with the experimental facts.

3.2. *Changes of Vibration Frequency*

An increase of frequency invariably reduces the spacing of the ridges. Thus, in one experiment using 220μ sand, frequencies of 440, 617, 867 and 1330 c/s obtained with plates of equal area and increasing thicknesses) resulted in distances between the ridges of 4·5, 3, 2·2 and 1·5 mm respectively. With a larger plate vibrating at 120 c/s, cork ridges were observed which were as much as 15 mm apart! Admittedly the amplitudes (like those of the diaphragm in Kundt's tube) decrease with increasing frequency, but, since intensity effects alone can only alter the spacing by a factor of two or three, this cannot wholly account for the above much larger variations in the spacing.

Table 3. Index to Figures 1–17

Figure	Subject	References
1	Granular, 14–18 grade (1100μ), ridges (sugar)	2.2
2	Powder, 120–160 (100μ), ridges (sand)	2.2, 2.7
3	Powder, 200–240 (70μ), ridges (sand)	2.7
4	Wavy ridges (lycopodium spores, 20μ)	2.7
5	Major and minor powder ridges	2.6
6	Major and minor granular ridges	2.6
7	Illustrating the new technique (sand)	1
8	Cylinders (spaghetti) in a double role	2.5, 2.6; table 2
9	Granular ridges, spheres (hundreds and thousands)	2.4 (*bii*); table 2
10	Granular ridges, cubes (table salt)	2.4 (*bii*); table 1
11	Granular ridges ellipsoids (Virginia stock seed)	2.4 (*bii*); table 2
12–17	Immovable objects (effects on ridges suggest presence of vortices)	2.5

3.3. *Distinction between Intensity and Frequency Effects*

From the last two paragraphs it appears that frequency effects are larger than, and to be distinguished from, intensity effects. Realizing that the frequency, as well as affecting the relative velocity, also governs the time which the viscous fluid is moving in the same direction, this is not to be wondered at. Thus, while the spacing of the ridges can be considerably increased by using larger intensities, wide spacing is essentially associated with larger particle diameters and lower frequencies. On the other hand, see again § 2.7 where ' waviness ' was discussed; close spacing is obtainable only at high frequencies and with appropriately smaller particles. At zero frequency, it may be worth noting, see Prandtl (1952, figure 3.22), the size of air circulations may assume enormous proportions. The senses of these are similar to Carrière's.

§ 4. CONCLUSIONS

From the experimental observations which have been described above, and the knowledge of vortex formations derived from the work of previous investigators, the following conclusions have been reached:

(i) Other conditions being constant, the distances between powder ridges increase roughly in proportion to the square root of the particle diameters. The ratio of spacing to diameter accordingly increases as the particles become smaller which provides evidence of variable shaped vortices (§§ 2.2, 2.3, 2.5).

(ii) ' Granular ' ridges (one particle wide) are more closely spaced than powder ridges which become thicker and further apart as the quantity of powder is increased (§ 2.2).

(iii) Elongated grains lie parallel to the ridge so that the specification of diameter in terms of sieve aperture is of considerable accuracy; height and breadth, not length, determine the spacing. Spheres touch in contrast to cubes which do not touch and set themselves diagonally. The main surface of flat particles is at right angles to the vibration vector (§ 2.4).

(iv) Density does not affect the ridge spacing appreciably. But it requires much larger intensities to form ridges with dense than with light materials. The movements with a vertical component of dense particles are less than, and different from, those of light particles (§ 2.2).

(v) At a given frequency, there is a lower limit of particle diameter below which straight ridges are no longer formed. They are replaced by sinusoidal and more widely spaced ridges, which however, disappear in favour of the original phenomenon of straight and closely spaced striations when the frequency of vibration is sufficiently raised (§ 2.7).

(vi) To form ridges with a given powder, a minimum intensity is required. By raising the intensity, the corresponding spacing can only be increased by a factor of two or three before major accumulations of the powdered material begins to take place (§ 3.1).

(vii) Intensity effects are thus smaller than, and to be distinguished from, frequency effects. It follows that the two variables which are primarily responsible for the ridge spacing are the particle diameter and the frequency. Close spacing is essentially associated with smaller diameters and higher frequencies (§§ 2.7, 3.3).

(viii) Minor ridges form their own independent vortices (§ 2.6).

(ix) A larger immovable object deflects the powder ridges in a manner which suggests the existence of vortices around it appropriate to its size. The powder patterns produced suggest a superficial analogy between hydrodynamic, and magnetic or electrostatic forces (§ 2.5).

(x) Some of the conclusions reached throw light on the dust striations obtained with electric sparks, a subject which it is hoped to consider at a later date. For the furtherance of knowledge concerning grain-fluid phenomena, experiments with graded powders and dusts in Kundt's tube, at sonic and ultrasonic frequencies are badly needed.

REFERENCES

ANDRADE, E. N., DA C., 1931, *Proc. Roy. Soc.*, A, **134**, 445; 1932, *Phil. Trans. Roy. Soc.* A, **230**, 417.

ANDRES, J. M., and INGARD, U., 1953, *J. Acoust. Soc. Amer.*, **25**, 928, 932.

BAGNOLD, R. A., 1946, *Proc. Roy. Soc.* A, **187**, 1.

CARRIÈRE, M. Z., 1929, *J. Phys. Radium*, **10**, 198.

COOK, S. R., 1902, *Phil. Mag.*, **3**, 471; 1903, *Ibid.*, **6**, 424.

DARWIN, G. H., 1883, *Proc. Roy. Soc.* A, **36**, 18.

FARADAY, M., 1831, *Phil. Trans. Roy. Soc.*, **121**, 299.

IRONS, E. J., 1929, *Phil. Mag.* (7), **7**, 523.

KUNDT, A., 1866 a, *Ann. Phys., Lpz.*, **127**, 497; 1866 b, *Ibid.*, **128**, 337.

KUNDT, A., and LEYMANN, O., 1874, *Ann. Phys., Lpz.*, **153**, 1.

PRANDTL, L., 1952, *Fluid Dynamics* (London: Blackie), ch. 3.

RAYLEIGH, LORD, 1884, *Phil. Trans. Roy. Soc.*, **175**, 1.

SKOGEN, N., 1952, *K. Norske Vidensk. Selsk. Forh.*, **24**, 60.

WALLER, M. D., 1954 a, *Nature, Lond.*, **174**, 368; 1954 b, *J. Sci. Instrum.*, **31**, 410.

RESEARCH NOTES

The Origin of Fine Structures in X-Ray Absorption Spectra of Ionic Solutions

BY AMAR NATH NIGAM
Physics Department, University of Allahabad

MS. received 13*th December* 1954, *and in revised form* 1*st February* 1955

§ 1. INTRODUCTION

THE presence of secondary structures in X-ray absorption spectra of ionic solutions was questioned for the first time by Cauchois (1948), and no satisfactory explanation seems to have been offered as yet. For solids Kronig's theory (1931) is found to fit in with the experimental facts, while for polyatomic gases another theory suggested by Kronig (1932) and developed by Petersen (1933) is found to hold good. In the case of an ionic solution, of simple salts (univalent, or divalent), no hypothesis has been advanced. In a solution, neither a crystal lattice nor the band system of a polyatomic molecule exists. The ions are supposed to be uniformly distributed in the solution and are in a state of Brownian movement. The object of this note is to suggest a probable cause.

§ 2. THE IONIC ATMOSPHERE AS A POTENTIAL WELL

The situation can be easily cleared if one considers a molecule or an ion to be confined for a short time in a potential space on account of attraction on all sides which prevents the ion from escaping. Consider for example the case of K^+ion in a solution of KCl in water. This ion is surrounded on all sides by an ionic atmosphere of Cl^- ions as suggested by Debye and Hückel (Taylor 1930). A photo-electron ejected by X-rays from a K^+ ion will find itself shut in a three-dimensional potential well formed by the Cl^- ions. The energies of this electron are similar to that of a particle in a rectangular potential box. The ejected electron will be repelled on all sides and the quantized energies of the electron are given by

$$E = \boldsymbol{h}^2 n^2 / 8\boldsymbol{m} d^2 \qquad \ldots\ldots(1)$$

where E is the energy of the band, n is an integer, $\boldsymbol{h}$ is Planck's constant, $\boldsymbol{m}$ the electronic mass and d the lattice constant. Due to spherical symmetry we have taken only one quantum number n. For solutions of good strength, the radius of the ionic sphere is the sum of the ionic radii of the two ions and therefore d in solutions will not differ much from d in crystal. The positions of maxima and minima in solutions will not differ appreciably from the corresponding positions in solids. The attached references show (Pauling 1929, Coiffari 1937, Ray Das and Bagchi 1940, Cauchois 1947, Setsuo Kiyono 1952, 1953) a collection of data for comparison. Intensity in case of solids is given (Smoluchowski 1935) by

$$I = H\,|V|^2 \qquad \ldots\ldots(2)$$

where H is a constant and V the potential. The corresponding equation for solutions can be written as

$$I = H\,|V/\mu|^2 \qquad \ldots\ldots(3)$$

where μ is the dielectric constant of the water molecule. The intensity distribution will thus depend on μ. This was observed by Vainshtein and Kavetskiĭ (1953).

Table 1. Comparison of Positions of the Structures in Solid and Solutions (Data in X.U.)

Substance	K-Edge	White line (r.b.)	Black line (r.n.)	Minimum	Maximum	Observer
Cl_2Ni—Solid	1483·13	1482·37	1481·3	1479·7	1477·5	Cauchois (1947)
Cl_2Ni—Solution	1483·13	1482·21	1480·7	1479·3	1477·7	
SO_4Ni—Solid	1482·95	1482·03	1480·9	1479·0	1477·2	
SO_4Ni—Solution	1482·98	1482·2	1480·9	1479·5	1477·8	
	K-Edge	A	α	B	β	
$KHCO_3$—Solid	3426·5	3425·0	3424·5	3423·8	3420·6	Kiyono (1952 & 1953)
$KHCO_3$—Solution	3427·3	3425·9	3424·2	3421·8	3416·0	
K_4FeCy_6—Solid	3427·1	3425·7	3424·1	3420·6	3415·9	
K_4FeCy_6—Solution	3426·9	3425·9	3423·5	3422·5	3420·3	
K_3FeCy_6—Solid	3426·8	3425·0	3423·5	3422·4	3420·6	
K_3FeCy_6—Solution	3427·2	3425·5	3424·0	3422·5	3421·4	
	K-Edge	a_1	k_1	a_2	k_2	
$Co(NO_3)_2$—Solid	1604	1598	1594	1592	1585	Ray, Das and Bagchi (1940)
$Co(NO_3)_2$—Solution	1603	1597	1594	1590	1586	
$Co(SO_4)$—Solid	1604	1598	1594	—	—	
$Co(SO_4)$—Solution	1604	1599	1595	1598	1589	
$CoCl_2$—Solid	1603	1600	1596	1593	1590	
$CoCl_2$—Solution	1603	—	—	1596	1591	

ACKNOWLEDGMENT

The author thanks Dr. G. B. Deodhar for his interest in the work.

REFERENCES

CAUCHOIS, Y., 1947, *C.R. Acad. Sci., Paris*, **224**, 1556; 1948, *Ibid.*, **227**, 65.
COIFFARI, B., 1937, *Phys. Rev.*, **51**, 634.
KRONIG, R. DE L., 1931, *Z. Phys.*, **75**, 191; 1932, *Ibid.*, **75**, 468.
PAULING, L., 1929, *Phys. Rev.*, **34**, 954.
PETERSEN, H., 1933 a, *Z. Phys.*, **76**, 768; 1933 b, *Ibid.*, **80**, 258.
RAY, B. B., DAS, S. R., and BAGCHI, N., 1940, *Indian J. Phys.*, **13**, 37.
SETSUO KIYONO, 1952, *Sci. Rep. Tôhoku Univ.*, series 1, **36**, No. 4, 236; 1953, *Ibid.*, series 1, **37**, No. 3, 249.
SMOLUCHOWSKI, R., 1935, *Z. Phys.*, **94**, 775.
TAYLOR, H. S., 1930, *Physical Chemistry*, Vo'. 1 (London: Macmillan), p. 784.
VAINSHTEIN, E. E., and KAVETSKIĬ, V. S., 1953, *C.R. Acad. Sci., U.R.S.S.*, **91**, 775.

LETTERS TO THE EDITOR

The Temperatures attained at the Surfaces of Growing Deposits, and the Origin of Stress in Deposits

Our recent electron diffraction and other results show that the temperatures attained at the surfaces of several types of deposits during their formation have hitherto been very much underestimated. We have found that under conditions normally used these may be several hundred degrees centigrade in deposits formed at about 15 to 300Å thickness per second, on metallic or other substrates initially at room temperature, by (i) condensation from the vapour in a vacuum, (ii) electrodeposition, or (iii) anodic oxidation. Our evidence appears conclusive and is as follows.

In condensed deposits of metals with high melting points, such as Cu, Ag, Au, Ni, Fe, Pd, in residual air at pressures of 10^{-5} to 10^{-6} mm Hg, the crystals first formed are small (<50 Å diameter) but the crystal size at the surface increases as deposition proceeds, and preferred orientation develops, associated with facet formation, in the upper parts of deposits more than 500–1000Å thick (Burgers and Dippel 1934, Evans and Wilman 1952). These observations all show the progressive large increase in mobility of the deposited atoms on the growing surface. This can only mean that the temperature at the deposit *surface* has substantially risen. Although Murbach and Wilman's (1953) substrate copper strips reached about 100° c and 150°c during deposition of Cu and Fe respectively under the conditions used, it is clear that the surface temperatures reached during deposition were much higher, higher in fact than (approximately) the normal recrystallizing temperature of the deposit metal, at which surface mobility first becomes appreciable, i.e. 200–230°c for Cu, 200°c for Ag and Au, 530–600°c for Ni, and 350–450°c for Fe.

It cannot be argued that the above increase in surface mobility is solely due to the temporary formation of a clean surface of deposited metal, associated with an amelioration of the vacuum, since (*a*) the vacuum was not so high initially as to allow the deposit surfaces to be regarded as abnormally clean, and Ag and Au are clearly not effective as ‘ getters ’ though the hot tungsten filament would have some such action, (*b*) it is known that Ag and Au recrystallize at a higher temperature in high vacuum than in presence of oxygen (see Finch and Fordham 1937).

Furthermore, theoretical conclusions that the surface temperature rise will be only a few degrees (e.g. Hoffman *et al.* 1954) are not valid because they assume that the surface is plane, which is quite incorrect (see Evans and Wilman 1952), and that the normal heat conductivity values apply as for solid metals. Evidence is recently available, from work elsewhere, of the abnormally low density of such metal deposits, their large surface area and very low electrical conductivity, all indicating a porous nature of the deposits, which is likely to be associated also with very low thermal conductivity.

In electrodeposits, metals such as Cu, Ag, Ni and Fe, deposited at rates similar to the above, develop relatively large (sometimes microscopic) crystals bounded by atomically smooth facets (Finch, Wilman and Yang 1947, Finch and

Layton 1951). The cathode surface can hardly be a strictly ' clean ' surface on which abnormally high mobility would be expected at room temperature; thus the logical conclusion is that here also the surface (at least to a few tens of atoms depth) has exceeded the recrystallizing temperature of the metal (see above). This is reasonable, since the energy liberated at the cathodic deposit surface amounts not only to the heats of condensation and solidification (to convert neutral metal atoms into a crystal), but also includes the six times larger energy of neutralization of the ions at the cathode, in place of the large influx of heat as radiation from the hot source of vapour in the condensed deposits discussed above (the energy of solvation of the ions is comparatively negligible).

Next, in anodic deposits, we have found (Kerr and Wilman 1955) that anodizing electropolished Be surfaces in an HNO_3–CrO_3 solution, at 31·5 mA cm^{-2}, yielded an oxide film which increased in thickness linearly with time ($\sim$80Å sec^{-1}) and which was practically amorphous ($<$10Å crystal diameter) after 15 seconds anodizing, but after 90 seconds was composed of crystals 60 Å in diameter at its surface. In view of our many observations that BeO either as thin films on the metal, or as isolated films, only recrystallizes appreciably at 300°c and higher, we interpret this as showing that the surface temperature well exceeded 300°c. In this case the energy liberated at the anode surface involves the neutralization of the O ions and the large heat of formation of the solid BeO, which is a poor conductor of heat. Anodic oxide layers on Al are normally amorphous or practically so, but occasionally the electron diffraction patterns show also rings due to crystals of γ-Al_2O_3 about 20Å in diameter, showing that at least parts of the surface exceeded 650°c, the recrystallizing temperature of such Al_2O_3. Hunter and Fowle (1954) recently obtained some evidence of a rise in surface temperature (to 124°c) of anodic Al_2O_3 layers, but only as an average over the surface and for particular conditions. Clearly it will be important to take account of the raised temperature of the oxide, in theoretical considerations of oxidation processes.

The above conclusions on condensed and also electrodeposited layers are fully supported by our successful explanation, in terms of the deposit growth at this raised temperature, of our observations on the stress developed in such deposits (Murbach and Wilman 1953)—its sign, magnitude, variation from one metal to another, and its dependence on the initial substrate temperature. The reduction and reversal of stress in Al in presence of oxygen (during condensation) was also paralleled by that caused by co-deposition of compounds during electrodeposition of Fe and Ni, while in absence of such disturbance tensile stresses of the same order as in condensed deposits were observed, again indicating a similar order of surface temperature.

Hoffman, Daniels and Crittenden (1954) hold another, less quantitatively supported view of the origin of the stress in condensed deposits, but their observations on decrease and subsequent rise in stress on vacuum-annealing the deposits are subject to the following serious doubts of interpretation: (*a*) mica substrates lose water from their surface layers at 200°c and above, with a 15% swelling of the lattice (Mutucumarana 1948) and have a probable oxidizing action on the deposits; (*b*) the transference of deposits to another apparatus for stress measurement, and the repeated reheating for 15–30 minutes *in vacuo* at 10^{-5}–10^{-6} mm Hg air pressure must have resulted in appreciable oxidation affecting the stress; (*c*) there is a slight but noticeable reaction of metals with mica (and also glass) at about 250°c and over, which could also affect the stress; (*d*) any oxide formed on or in the deposit at the lower temperatures of annealing would be amorphous and would crystallize on raising the temperature sufficiently, e.g. Cu_2O at 130–140°c (Leu 1951), NiO at about 150°c (Sinha 1954), thus affecting the stress. They also misread our paper—we stated clearly that we considered only the immediate surface region of the deposits to attain these high temperatures.

Chemical Engineering Department,
Imperial College, London, S.W.7.
5th March 1955.

H. WILMAN

BURGERS, W. G., and DIPPEL, C. J., 1934, *Physica*, **1**, 549.
EVANS, D. M., and WILMAN, H., 1952, *Acta Cryst.*, *Camb.*, **5**, 731.
FINCH, G. I., and FORDHAM, S., 1937, *J. Soc. Chem. Ind.*, *Lond.*, **56**, 632.
FINCH, G. I., and LAYTON, D. N., 1951, *J. Electrodep. Tech. Soc.*, **27**, 215.
FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Disc. Faraday Soc.*, **1**, 144.
HOFFMAN, R. W., DANIELS, R. D., and CRITTENDEN, E. C., 1954, *Proc. Phys. Soc.* B, **67**, 497.
HUNTER, M. S., and FOWLE, P. F., 1954, *Trans. Electrochem. Soc.*, **101**, 514.
KERR, I. S., and WILMAN, H., 1955, in course of publication.
LEU, K., 1950, *Chimia*, **4**, 263; 1951, *Thesis*, University of London.
MURBACH, H. P., and WILMAN, H., 1953, *Proc. Phys. Soc.* B, **66**, 905.
MUTUCUMARANA, T. DE S., 1948, *Thesis*, University of London.
MUTUCUMARANA, T. DE S., and WILMAN, H., 1954, paper read at 3rd International Congress of Crystallography, Paris, (see 1954, *Acta Cryst.*, *Camb.*, **7**, 666. (A)).
SINHA, A. P. B., 1954, *Thesis*, University of London.

REVIEWS OF BOOKS

Physics of Particle Size Analysis. A conference arranged by the Institute of Physics and held in the University of Nottingham from 6th to 9th April 1954. Pp. xii + 218. (London: Institute of Physics, British Journal of Applied Physics, Supplement No. 3, 1954). 35*s*.

The third supplement to the British Journal of Applied Physics contains 33 papers, three surveys and a summary presented at a Conference on the physics of particle size analysis held at Nottingham University in April, 1954. Seven years earlier a conference on the same subject took place in London, but the recent one was on a far more ambitious scale and covered the fundamental scientific aspects in a more comprehensive manner.

Most of the papers came from Universities or Government research establishments, with specialized industrial applications appearing in only a few. The coal industry was strongly represented, but the work described had implications beyond the field in which it had been conceived.

The Conference was divided into eight independent sessions which were held consecutively. The first two dealt with the relative motion of particles and fluid, session A covering fluid mechanics and session B molecular phenomena. Apart from a paper on viscosity the former was devoted to sedimentation, two papers being about the paths of particles which settled in moving air and one about the fall of aggregates.

In session B three papers treated gas flow through beds of particles and estimation of the surface area of fine powders by this method, with due allowance for adsorption when the flow is not steady. There was also an interesting study of a pseudo-viscous effect produced by the reaction upon charged particles settling through an electrolyte of a potential gradient set up by their motion.

A session on light scattering followed with a survey and two papers. One showed that selective sampling is really needed for optical analysis of clouds containing particles of different sizes because the size-dependent effects produced by the smaller particles are swamped by light scattered from the large. The other paper confirmed by experiments with uniform suspensions that the scattering coefficient can be expressed as a unique function of $4\pi r(m-1)/\lambda$ where r is the particle size, m their refractive index and λ the wavelength of light.

Session D was concerned with shape factors. Microscopic sizing is inevitably referred to the projected diameter while Stokes' diameters are required for comparison with sedimentation. The relationship between the diameters was analysed and it was shown to be necessary to take thickness into account as well as the projected size. One author had a microscopic method of measuring thickness directly.

Particles which are counted under a microscope within a certain range of projected diameter will cover a wider range of Stokes' diameter because of the distribution of shape factor. It is shown how size frequency distribution in Stokes' diameters, mass or surface area can be computed from observations in projected diameters.

Classification by rate of fall, both in gravitational and centrifugal fields, revealed that particles of dust falling at a given speed covered a wide range of projected size. Results were expressed in terms of shape factor or effective density of particle. The falling speed depends on density, shape and orientation as well as size.

Session E consisted of four papers on visual counting and sizing. Interlaboratory comparisons revealed the need for improved mounting and microscopic technique as well as prior agreement about aggregates. For sampling in connection with coal pneumoconiosis aggregates should be reckoned as single particles, but care must be taken to see that the number of particles on the slide is not so great as to cause overlapping. A new method of securing dust samples for the electron microscope was described and comparison counts with this instrument and the visual one were analysed.

Sessions F and G dealt respectively with the theory and practice of counting and sizing particles automatically. Seven working systems were described in the latter, using principles which included the registration of individual particles swept past a recorder one at a time and track or line scanning of either deposited particles or photographs. Ingenious solutions involving full exploitation of the power of electronics have been proposed for recording size distribution and difficult problems involving mechanical movements of great precision and the stability of light sources have been successfully solved. The full resolving power of the visual microscope has not yet been attained by these automatic devices.

The final session contained papers on methods of mounting, the adhesion of dust particles and the aggregation of aerosols. C. N. DAVIES.

Advanced Level Examples in Physics, by R. V. Johns and W. F. Ware. Pp. viii + 346. (London: Macmillan, 1954.) 16*s*.

The authors largely succeed in their aim of supplementing the theoretical and laboratory work of students preparing for examinations in Physics at the Advanced Level of the G.C.E. and for Intermediate examinations, particularly in the case of the weaker candidates.

The book contains about one thousand examples of graded difficulty, arranged by topics, with answers at the end. Each section is preceded by a useful summary of relevant principles and formulae, together with worked examples. These examples are very lucidly set out. In the main, arithmetical methods are employed in preference to those involving juggling with formulae, and ample verbal explanation of the steps is given. In a few cases only, unnecessarily long methods of solution are given.

One or two omissions have been noticed. There are not many examples involving the wave theory of light, and there is no mention of this subject in the text, nor are there any worked examples upon it. The sections on electromagnetism seem rather curtailed: for example, self induction is not mentioned, and there is but a bare minimum on alternating current, inadequate for the Advanced Level syllabus of at least one examining body.

There is a useful set of over one hundred revision examples at the end of the book, and many of these are suitable for the brighter students.

The 'real is positive' optical sign convention is used.

R. SHEWELL.

BOOK NOTICES

Science and its Background, by H. D. ANTHONY. 2nd Edn. Pp. ix + 337. (London: Macmillan, 1954.) 20*s.*

The Sources of Eddington's Philosophy, by H. DINGLE. Pp. vi + 64. (Cambridge: University Press, 1954.) 3*s.* 6*d.*

Équilibres physico-chemiques et données thermodynamiques des mélanges gazeux aux températures élevées, by G. RIBAUD and N. MANSON. Pp. iii + 111. (+29 tables). (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 294, 1954.) 1,200 *fr.*

Étude de la déformation et de la rupture des matières plastiques, by F. ZANDMAN. Pp. iv + 198. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 291, 1954.) 2,300 *fr.*

Luminescence—with particular reference to inorganic phosphors. A Symposium held by the Electronics Group of the Institute of Physics at the Cavendish Laboratory, Cambridge, from 7th to 10th April, 1954. Pp. iv + 120. (London: Institute of Physics, 1955.) British Journal of Applied Physics Supplement No. 4. 25*s.*

Equazioni alle derivate parziali di tipo ellittico, by C. MIRANDA. Pp. viii + 222. (Berlin, Göttingen, Heidelberg: Springer, 1955.) (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Heft 2.) DM 28.80.

Transfinite Zahlen, by H. BACHMANN. Pp. vii + 204. (Berlin, Göttingen, Heidelberg: Springer, 1955.) (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Heft 1.) DM 29.80.

Analytische Fortsetzung, by L. BIEBERBACH. Pp. 167. (Berlin, Göttingen, Heidelberg: Springer, 1955.) (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Heft 3.) DM 24.80.

Mécanique ondulatoire et cinétique chimique. Réunions d'études et de mises au point, tennues sous le présidence de Louis de Broglie. Pp. 204. (Paris: Éditions de la Revue d'Optique Théorique et Instrumentale, 1955.) 1500 *fr.*

Tuyères supersoniques à col réglable, by M. MÉNARD and O. MONOD. Pp. ii + 33. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, 1955.) (Notes Techniques, No. 51.) 400 *fr.*

Chaleurs spécifiques et entropies des hydrocarbures et de quelques combustibles liquides, by M. AUBERT and N. SIVOLOBOV. Pp. iv + 202. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 297, 1955.) 1900 *fr.*

CONTENTS OF SECTION A

PAGE

Mr. P. C. Price. The Radiative Capture of Alpha Particles in ^{14}N 553

Dr. G. N. Fowler. On the Scattering of High Energy Electrons 559

Mr. J. T. Lewis, Mr. M. R. C. McDowell and Dr. B. L. Moiseiwitsch. Properties of the Hydrogen Molecular Ion: V—Transitions Connecting the Lowest Even and Lowest Odd π-States with Higher σ-States 565

Mr. T. R. Carson and Dr. A. Dalgarno. Approximate Molecular Orbitals: III—The 2sσ_4 and 3dσ_4 states of H_2^+. 569

Prof. H. S. Green. Covariant Treatment of the Nucleon–Nucleon Interaction . 577

Prof. C. B. A. McCusker, Mr. J. G. Dardis and Mr. B. G. Wilson. The Rates of Penetrating Showers at Sea Level 585

Prof. H. C. Longuet-Higgins and Dr. J. A. Pople. The Electronic Spectra of Aromatic Molecules: IV—Excited States of Odd Alternant Hydrocarbon Radicals and Ions 591

Prof. H. C. Longuet-Higgins and Mr. J. N. Murrell. The Electronic Spectra of Aromatic Molecules: V—The Interaction of Two Conjugated Systems . . 601

Dr. M. A. S. Ross, Mr. A. J. Cochran, Dr. J. Hughes and Prof. N. Feather. The Fluorescence Yields of the L Levels of Bismuth 612

Mr. H. Trefall. On the Positive Temperature Effect in the Cosmic Radiation . 625

Dr. J. T. Lewis. Ionic Configuration Interaction in Some Excited States of the Hydrogen Molecule 632

Mr. A. Ward and Dr. P. J. Grant. The Reaction ^{11}B(d, n)^{12}C 637

Research Notes:

- Dr. B. Donovan, Dr. A. B. Lidiard and Dr. N. H. March. On the Spin Paramagnetism of a Free Electron Gas: II 644
- Mr. W. A. Runciman. Absorption and Emission Spectra of Bismuth-Activated Phosphors 647
- Dr. W. D. Allen, Dr. A. T. G. Ferguson and Mr. J. Roberts. Neutron–Deuteron Scattering in the Range 80–200 kev 650
- Mr. B. Nordfors. A Note on the Al K$\alpha_3\alpha_4$ Lines in Metal and Oxide . . 654
- Dr. Satio Hayakawa and Dr. Shiro Yoshida. Inelastic Scattering by Rotational Excitation 656

Reviews of Books 659

Contents of Section B 664

Figure 1. ×0·9

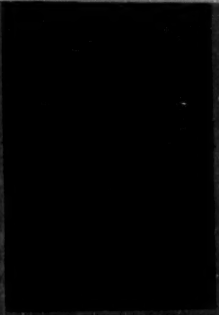


Figure 2. ×0·9

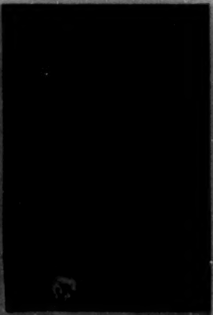


Figure 3. ×0·9

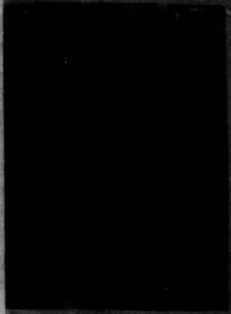


Figure 4. ×0·45

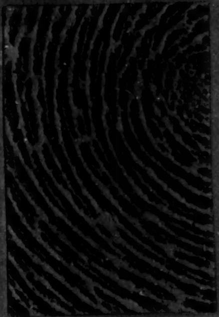


Figure 5. ×0·9

Figure 6. ×2·5

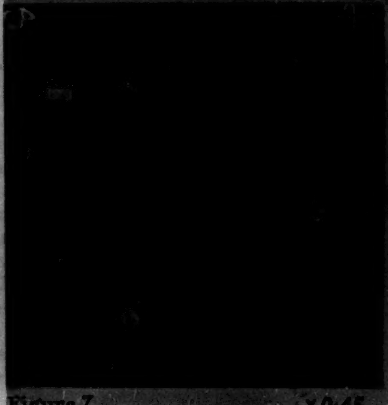


Figure 7. ×0·45

Figure 8. ×0·8

PLATE I.

Figure 9.

Figure 10.

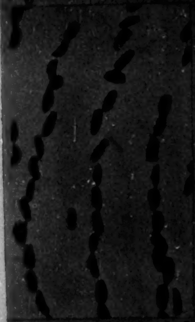


Figure 11. ×0·6

[illegible]

Figure 14. ×0·7

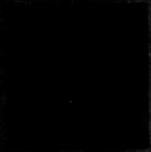


Figure [illegible] ×0·7

Figure 15. ×0·7

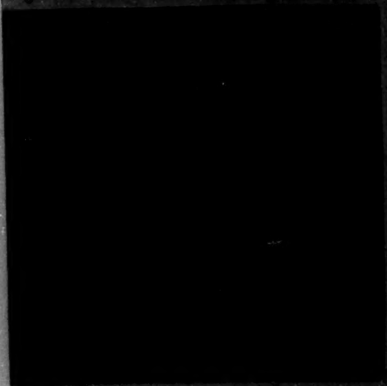


Figure 16. ×0·[illegible]

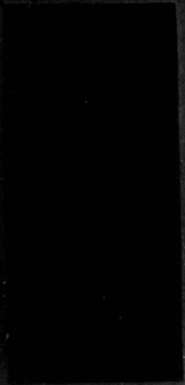


Figure 17. ×0·7

PLATE II.

The Effects of Anisometric Amplitude Patterns in the Measurement of Ionospheric Drifts

BY G. J. PHILLIPS† AND M. SPENCER
Cavendish Laboratory, Cambridge

Communicated by J. A. Ratcliffe ; MS. received 28th December 1954
and in final form 17th March 1955

Abstract. In certain observations of a downcoming radio wave, made for the purpose of measuring horizontal drifts of ionospheric irregularities, time shifts between fading curves obtained at three spaced receiving points are used to calculate the velocity and direction of drift of the amplitude pattern over the ground. Two simplifying assumptions about the pattern are frequently made, namely (i) the detail remains constant as it moves, and (ii) it is statistically isometric, i.e. the separation of two points for which the amplitudes have a given fixed correlation is independent of direction.

The errors that arise if assumption (i) is wrong have been considered in a published paper. But a more general analysis, not relying on either assumption may be applied to the three fading curves, and the result expressed (either analytically or graphically) as the error of a calculation which makes the two assumptions. One important result is that if assumption (ii) is not valid the error in direction of drift may be quite serious. Thus, in an application to amplitude patterns produced by vertical reflection at 2·4 Mc/s (principally from the E layer), errors in individual results averaged 15°. Because the structure of the pattern varied from sample to sample, the error was not, in this case, systematic. But, in some applications, a preferred direction of the *pattern structure* could, if ignored, lead to wrong conclusions about preferred directions of *drift*.

§ 1. INTRODUCTION

Two methods of measuring horizontal drifts of irregularities of ionization in the ionosphere make use of the irregular amplitude pattern produced on the ground by a downcoming radio wave. One method uses waves returned by the ionosphere from a sender on the ground and the other uses radiation transmitted through the ionosphere from one of the more intense radio stars. A horizontal movement of the ionospheric irregularities in either case causes the amplitude pattern to drift over the ground.

In most of the experimental work using reflected waves, for example that of Mitra (1949), Krautkrämer (1950), Salzberg and Greenstone (1951) and Phillips (1952), complete fading curves were recorded on photographic film or by fast-running pen recorders. Pulse transmissions in the frequency range 2–6 Mc/s were generally used, with three receiving aerials placed at the corners of a right-angled triangle of 100–200 m side and situated near the transmitter. Three records of fading were thus obtained, and the time shifts between pairs of records were estimated. The drift of the pattern was then deduced by a simple calculation, based on the assumptions that the pattern was (i) constant in detail as it moved, and (ii) statistically isometric, i.e. its statistical properties showed circular symmetry.†

---

† Now with the British Broadcasting Corporation.

This paper is primarily concerned with what happens when the pattern is statistically anisometric, so that the contours of constant auto-correlation are not circular. It is shown that, under some conditions, the deductions made on the simple assumption might be seriously in error, and it is shown how the conditions can be recognized. If the amplitude pattern changes as it drifts over the ground, additional complications arise. Briggs, Phillips and Shinn (1950) have shown how to deal with these if the pattern is isometric. In Appendix II of the present paper the corresponding problem for an anisometric pattern is discussed.

An extreme case will illustrate the error which arises when a pattern is wrongly assumed to be isometric. Consider for example a pattern in which the amplitude contours are so elongated that they appear as a set of parallel lines. It is clear that motion parallel to the lines can never be detected, and only the component of motion perpendicular to them will be measured. When the pattern has a finite structure in all directions the situation is not so serious, but errors may still arise if it is not completely isometric.

We shall not consider in this paper the relation between drifts of the pattern and drifts in the ionosphere; this topic is dealt with in the references cited above.

§ 2. STATISTICAL PROPERTIES OF AN ANISOMETRIC PATTERN

This section will be concerned with the nature of the amplitude pattern on the ground at any instant; the determination of drifts is not considered until § 3.

2.1. *The Representation of a Statistically Anisometric Pattern*

Figure 1 (*a*) shows the contours of constant amplitude of a pattern of the type we shall consider. Figure 1 (*b*) represents its two-dimensional auto-correlation function $\rho(\xi, \eta)$, i.e. the correlation over many samples between the amplitudes at two points of the type (x, y) and $(x+\xi, y+\eta)$ in a cartesian coordinate system.

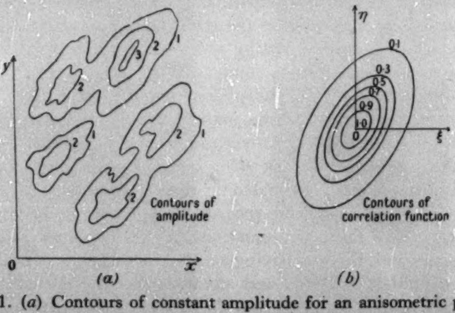


Figure 1. (*a*) Contours of constant amplitude for an anisometric pattern.
(*b*) Contours of equal auto-correlation for the same pattern.

---

† Briggs, Phillips and Shinn (1950) used the word 'isotropic' to describe this property.

If a pattern is not statistically isometric, it is likely in many practical cases that the contours of constant correlation will take the form of similar concentric ellipses as shown in figure 1 (*b*). A pattern of the type shown in figure 1 (*a*) could be produced from a statistically isometric pattern, for which the correlation contours would be circular, by stretching it in one direction. Distortion of this type may be represented by a 'characteristic ellipse', similar to the contours of constant correlation, which may be said to give an 'average shape' of the contours of constant amplitude in figure 1 (*a*). Given (i) the axial ratio and (ii) the orientation of the ellipse, together with (iii) the auto-correlation function for only one direction along the ground, the complete auto-correlation of figure 1 (*b*) may be deduced. If its contour-shape only is required, it is sufficient to determine (i) and (ii).

2.2. *Practical Determination of the Characteristic Ellipse*

Ideally, by observation at a large number of points, the statistics of the pattern of amplitude on the ground could be investigated without reference to changes with time or to drifting. In practice, however, it is these effects which serve to bring various samples of the pattern to a limited number of fixed observing points. Therefore the analysis of three-point fading curves is first considered for the purpose of finding the parameters of the characteristic ellipse, and the question of drifts is left until the next section.

Let $\rho(\xi, \eta, \tau)$ denote the correlation between amplitudes recorded at two points with relative displacement (ξ, η) in cartesian coordinates, and at times separated by τ†. The function $\rho(\xi, \eta)$ of § 2.1 may now be obtained as $\rho(\xi, \eta, 0)$ by direct cross-correlation of a pair of fading records. The form which we have assumed for this function (figure 1 (*b*)) is such that a plot against spacing $\sqrt{(\xi^2+\eta^2)}$ along any direction on the ground (space correlogram) has a similar *shape* though varying in scale. We shall assume in the following that the shape is also similar to that of the time correlogram $\rho(0, 0, \tau)$.

Consider the analysis of fading curves for points $(0, 0)$, $(a, 0)$ and $(0, a)$ at the corners of an isosceles right-angled triangle of side a. We proceed thus:

(i) From one fading curve determine $\rho(0, 0, \tau)$ for a number of values of τ; by plotting the results against τ the time correlogram is obtained. (Since this function is the same for any point on the ground, only one fading curve need be used.)

(ii) Determine $\rho(a, 0, 0)$, $\rho(0, a, 0)$ and $\rho(a, -a, 0)$ from a direct cross-correlation analysis between pairs of curves.

(iii) From the time correlogram find the values of τ, namely τ_1, τ_2 and τ_3, which give values of ρ equal, respectively, to the three values determined in (ii).

Stage (iii) above has converted the measurements of cross-correlation into 'equivalent time shifts'. Thus we find that amplitudes at points displaced distance a along Ox are correlated to the same extent as amplitudes at time τ_1 apart on a single fading record. This would occur, for example, if the pattern were drifting bodily parallel to Ox with a velocity $(V_c')_{0x} = a/\tau_1$; in general this is not the case, but the ratio defined is still a useful concept called by Briggs, Phillips and Shinn (1950) the 'fading velocity'. It gives the factor by which the τ scale of the time correlogram must be multiplied in order that it should represent the space correlogram for the direction considered. In our example, V_c' takes values a/τ_1, a/τ_2 and $\sqrt{2}a/\tau_3$ for the three base-line directions. The simplest example of spacings in these directions which have a fixed correlation is given by putting $\tau=1$; the spacings are then equal to V_c', the correlation being equal to $\rho(0, 0, 1)$. It follows that a polar plot of V_c' must be one of the ellipses of constant correlation considered in § 2.1. The known values of V_c' give the length of semi-diameters of the ellipse for only three directions, but this is sufficient to determine the axial ratio $r (r>1)$ and orientation ψ of the minor axis (measured anticlockwise from Ox). This stage is summarized below, the necessary formulae being derived in Appendix I.

---

† A formal definition of this correlation function is given by Briggs, Phillips and Shinn (1950). However, throughout the paper, certain other functions which vary with correlation may be used in its place. Thus the 'after-effect function' Δ (see Fürth and MacDonald 1947) which is the average magnitude of the difference between normalized fading records, is very convenient, being easier to evaluate—by graphical or other means—than ρ. A minimum of Δ, of course, indicates a maximum of ρ.

(iv) Calculate r and ψ from the expressions

$$r^2=\frac{R+(P^2+Q^2)^{1/2}}{R-(P^2+Q^2)^{1/2}} \quad \text{and} \quad \tan 2\psi = P/Q,$$

where $P=\tau_1{}^2+\tau_2{}^2-\tau_3{}^2$, $Q=\tau_1{}^2-\tau_2{}^2$ and $R=\tau_1{}^2+\tau_2{}^2$. Ambiguity of ψ in the range $0 \leqslant \psi < 180°$ is resolved by taking the value giving $\cos 2\psi$ of the same sign as Q.

§ 3. DETERMINATION OF THE DRIFT OF AN AMPLITUDE PATTERN

3.1. *Theoretical Results*

Drift of an amplitude pattern is indicated by a significant average time shift between one or more pairs of records of fading, such as those taken at three points at the corners of a right-angled triangle. The average time shift between two records may be defined as that time displacement which gives rise to the maximum correlation between them.

If the amplitude pattern is assumed to be statistically isometric, and to move without changing its form, the velocity V and direction of drift ϕ (measured anticlockwise from Ox) may be deduced by the equations given by Briggs, Phillips and Shinn (1950), giving

$$(1/V)^2=(1/V_1')^2+(1/V_2')^2$$

and

$$\tan\phi = V_1'/V_2'.$$

The equations represent the solution of equations (A) and (B) in Appendix II. V_1' and V_2' are the ratios of receiving aerial separation to average time shift for pairs of points lying along axes Ox and Oy respectively.

If we use these equations when the assumptions are not justified, we shall determine apparent values of velocity and direction which we may denote by V_a and ϕ_a. The complete solution of the problem when the pattern is anisometric and changing in detail as it moves is given in Appendix II; the results obtained are given in the following two equations which, it will be noted, do not involve the orientation of the base line of the receiving points:

$$\tan(\phi-\phi_a)=\frac{(r^2-1)\tan(\phi_a-\psi)}{1+r^2\tan^2(\phi_a-\psi)} \qquad \ldots\ldots(1)$$

and

$$V_a/V'=\cos(\phi-\phi_a). \qquad \ldots\ldots(2)$$

Here r and ψ are the axial ratio and orientation of the V_c' ellipse as already defined.

V' is a quantity related to the true velocity V by equation (M) in Appendix II; as we are now principally concerned with the errors arising when the patterns are anisometric, it is sufficient to remark here that V' is equivalent to the true velocity of the pattern when it is possible to neglect changes in detail of the pattern as it moves.

The difference $\phi-\phi_a$ between the true and apparent directions of drift is related to the angle $\phi_a-\psi$ between the apparent direction and the minor axis of the ellipse as shown in figure 2. It depends on r, the axial ratio of the ellipse. It will be seen that the difference may be large, e.g. values up to 55° can occur when $r=3$. The relation between the ratio V_a/V' and the angle $\phi_a-\psi$ is shown in figure 3. There are no errors in velocity or direction when the minor or major axis of the ellipse is directed along the true direction of drift.

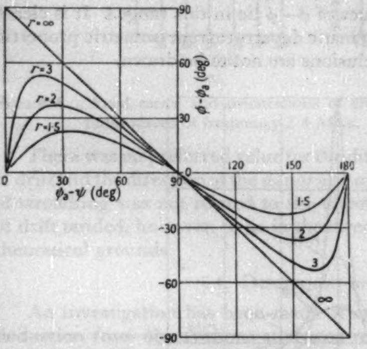


Figure 2. The required correction $\phi-\phi_a$ plotted as a function of the angle $\phi_a-\psi$ between the apparent direction and the minor axis of the charateristic ellipse, for various axial ratios of the ellipse.

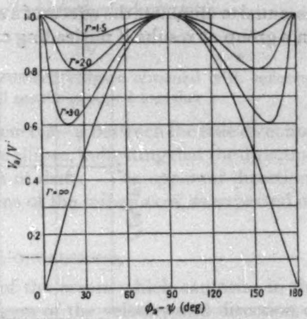


Figure 3. The ratio of V_a, the apparent velocity, to V', the velocity which would be deduced from a correct analysis, plotted in the same way as figure 2.

3.2. *The 'Grouping' of Apparent Directions*

In this section we shall investigate the effect of a systematic departure from isometric properties, viz. the case where r and ψ are constant. Consider a group of results in which ϕ takes all values equally. The apparent directions ϕ_a will then be crowded together about the most probable value, ψ. This may be seen by inspection of the curves of figure 2; there is always a tendency for the apparent direction to be nearer to direction ψ than the true direction is.

The probability distribution of apparent directions in a group of results may be found as follows. The proportion of values between ϕ and $\phi+d\phi$ (with corresponding apparent directions between ϕ_a and $\phi_a+d\phi_a$) may be written, for ψ constant, as

$$p(\phi-\psi)d(\phi-\psi)=p(\phi_a-\psi)d(\phi_a-\psi).$$

Differentiating equation (F), Appendix II, we have

$$[1+\tan^2(\phi-\psi)]d(\phi-\psi)=r^2[1+\tan^2(\phi_a-\psi)]d(\phi_a-\psi).$$

If $\phi-\psi$ takes all values equally between 0 and π (there is no distinction between $\phi-\psi$ and $\pi+\phi-\psi$), we may write $p(\phi-\psi)=1/\pi$. From the above equations and equation (F) we then obtain the distribution of $\phi_a-\psi$ as

$$p(\phi_a-\psi)=p(\phi-\psi)\,\frac{d(\phi-\psi)}{d(\phi_a-\psi)}=\frac{[1+\tan^2(\phi_a-\psi)]/\pi}{r^{-2}+r^2\tan^2(\phi_a-\psi)}.$$

The most likely value is $\phi_a-\psi=0$, for which $p_{\max}(\phi_a-\psi)=r^2/\pi$.

Distribution curves for various values of r are shown in figure 4. They illustrate the extent to which the values of apparent direction of drift ϕ_a cluster round the direction ψ of the minor axis of the ellipse; the distribution of $\phi-\psi$ is the same as the curve for $r=1$. As r increases, the effect becomes increasingly marked. When $r=3$, for example, 93% of the values of $\phi_a-\psi$ are contained in the range $-45°$ to $+45°$, whereas on our assumption that the values of ϕ were uniformly distributed, 50% of the values of $\phi-\psi$ lie in this range. It is clearly essential to eliminate the effects of a systematic departure from isometric properties in a group of results if misleading conclusions are not to be drawn.

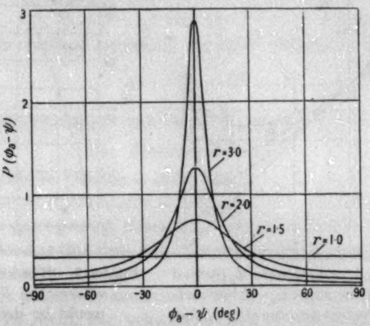
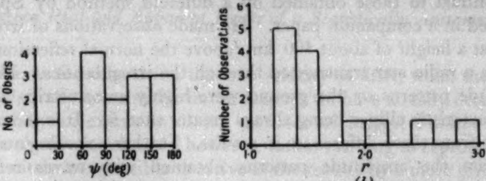


Figure 4. The grouping of apparent directions for various departures from isometric properties assuming a uniform distribution of true directions.

3.3. *Experimental Results*

Some records of the fading of radio waves of frequency 2·4 Mc/s reflected from the ionosphere at vertical incidence have been analysed to determine the importance of anisometric properties of the amplitude patterns. Samples approximately 10 fading cycles long were taken from recordings made on 15 different occasions. Ten were obtained with reflections from the E region, two from the sporadic E region, and three from the F region. The amplitude pattern was found to be statistically anisometric in many cases, but the results deduced from different records were not the same. The values of ψ were about equally distributed through 180° figure 5 (*a*), showing no preferred direction for the axes of the ellipses. The largest value of r was 2·9, obtained with reflection from the E region, and the most frequently occurring was 1·3 (figure 5 (*b*)). The largest values for E_s and F reflections were 2·2 and 1·6 respectively. A rough test showed that values of this order have a real statistical significance for the lengths of record analysed.

The errors in V and ϕ arising from the use of formulae appropriate to isometric patterns were evaluated for each sample. The greatest error in direction was 53°, and the mean error 15°. The mean value of V_a/V' was 0·94, and the smallest was 0·60. From these results it appears that the errors caused by anisometric patterns were not, on average, serious, though individual results could be considerably in error. Erroneous conclusions might have been drawn if there had been a preferred value for ψ, since the apparent directions would then have been grouped around the direction corresponding to the most probable value of ψ.



(a) *(b)*

Figure 5. Axial ratios and orientations of characteristic ellipses obtained with reflected radio waves of frequency 2·4 Mc/s. (All results grouped together.)

There was no preferred value for the difference $\phi - \psi$ between the true direction of drift and the direction of the minor axis of the ellipse, indicating that the direction of stretching was not related to the direction of drift. The *apparent* directions of drift tended, however, to lie in the directions of the minor axes, as expected on theoretical grounds.

§ 4. DISCUSSION AND CONCLUSIONS

An investigation has been made of one of the errors which can arise in the deduction from observations at spaced receivers of the velocity and direction of drift of an amplitude pattern. The error arises when it is wrongly assumed that the pattern is statistically isometric, i.e. that the irregularities in it have the same average size in all directions. A less stringent assumption has been made, namely that the pattern is one that may be produced by the uniform stretching of a statistically isometric pattern, so that its properties may be represented by the 'characteristic ellipse' which would be formed on stretching a circle in the same manner.

The errors in the magnitude and direction of the measured velocity depend on the orientation of the ellipse and on its axial ratio, as shown in figures 2 and 3. No errors arise when one of the axes of the ellipse is parallel to the true direction of drift. The errors may be serious for other orientations. If the pattern is always stretched in the same direction in a group of results, the apparent directions of drift will tend to group themselves about the direction of the minor axis of the characteristic ellipse.

Experimental investigation of the amplitude patterns obtained with waves reflected from the ionosphere at vertical incidence has shown that they are often statistically anisometric. In the experiments a frequency of 2·4 Mc/s was used and, for reflections from the E layer, sufficient samples were taken under similar conditions but at widely spaced intervals of time to show that the characteristic ellipse was variously orientated. It therefore appears that the mean of a number of velocities deduced from this type of experiment should not be subject to a systematic error. The errors due to anisometric patterns will, however, contribute to the considerable fluctuation in half-hourly mean values of the magnitude and direction of velocity which are observed by workers in this field. It seems permissible to use the simple 'isometric' formulae for routine measurements of this kind, provided that the possibility of errors in individual results is appreciated.

For reflections from the F layer preliminary results (only three examples) do not suggest any marked difference from the results for the E layer. This is in striking contrast to those obtained by a different method by Spencer (1955) and reported in a companion paper. He made observations of irregularities of ionization at a height of about 400 km (above the normal reflection level) using waves from a radio star transmitted through the irregularities. He found that the amplitude patterns on the ground were highly anisometric, the axial ratio of the characteristic ellipse being always greater than 5. It is too early to draw conclusions from these differences in results. Further experiments are needed to investigate the amplitude patterns obtained with waves reflected from different heights within the F region.

The experimental results for reflected waves have not shown any relation between the orientation of the characteristic ellipse and the true direction of drift.

APPENDIX I

The Constants of the Characteristic Ellipse

The equation of an ellipse of axial ratio r $(r>1)$, referred to axes x_1, y_1 along the minor and major axes may be written

$$r^2x_1^2+y_1^2=c^2.$$

For axes rotated clockwise by an angle ψ, the equation becomes

$$x^2(r^2\cos^2\psi+\sin^2\psi)+2xy(r^2-1)\sin\psi\cos\psi+y^2(\cos^2\psi+r^2\sin^2\psi)=c^2.$$

Let the intercepts on $y=0$, $x=0$ and $x=-y$ be u, v and $\sqrt{2}w$ respectively. Then

$$u^2(r^2\cos^2\psi+\sin^2\psi)=c^2, \qquad v^2(\cos^2\psi+r^2\sin^2\psi)=c^2$$

and
$$w^2[r^2+1-2(r^2-1)\sin\psi\cos\psi]=c^2.$$

If the ellipse represents a polar graph of V_c', we shall have

$$u=a/\tau_1, \qquad v=a/\tau_2 \qquad \text{and} \qquad \sqrt{2}w=\sqrt{2}a/\tau_3,$$

where τ_1, τ_2 and τ_3 are the equivalent time shifts defined in § 2.2. Hence, from above,

$$a^2(r^2\cos^2\psi+\sin^2\psi)/\tau_1^2=c^2, \qquad a^2(\cos^2\psi+r^2\sin^2\psi)/\tau_2^2=c^2$$

and
$$a^2[r^2+1-2(r^2-1)\sin\psi\cos\psi]/\tau_3^2=c^2.$$

Consider
$$P=\tau_1^2+\tau_2^2-\tau_3^2=a^2(r^2-1)\sin(2\psi)/c^2$$
$$Q=\tau_1^2-\tau_2^2 \qquad =a^2(r^2-1)\cos(2\psi)/c^2$$

and
$$R=\tau_1^2+\tau_2^2 \qquad =a^2(r^2+1)/c^2.$$

From these relations we have $\tan 2\psi=P/Q$

and
$$(r^2-1)/(r^2+1)=(P^2+Q^2)^{1/2}/R$$

or
$$r^2=[R+(P^2+Q^2)^{1/2}]/[R-(P^2+Q^2)^{1/2}].$$

The ellipse may also be built up graphically from the three values u, v and $\sqrt{2}w$, using the fact that the diameter of an ellipse bisecting a chord bisects all chords parallel to it. This method is no more difficult to apply when the aerial spacings are not equal, or when the base lines are not at right angles, whereas the algebraic formulae would become more complicated.

APPENDIX II

Relations between Real and Apparent Drifts of a Statistically Anisometric Pattern

(a) Methods.

The equations for calculating the true velocity of drift, when the amplitude pattern is statistically anisometric in the manner we have already described, may be established by an extension of the methods of Briggs, Phillips and Shinn (1950). The algebra involved is, however, rather too long to reproduce in detail here. Fortunately it is possible to arrive at the results by a more direct method. The results for a statistically isometric pattern are first summarized in a suitable geometrical form, and then adapted to apply to the case of an anisometric pattern. Once a geometrical construction has been established to give the velocity of drift in this more general case, it is a simple matter to obtain the corresponding algebraic formulae.

(b) Summary of results for a statistically isometric pattern.

The basic results for an isometric pattern which does *not* change in detail as it moves may be given as

$$V_n = V\cos\phi_n \qquad \text{......(A)}$$

and

$$V = V_n'\cos\phi_n. \qquad \text{......(B)}$$

V is the velocity of drift of the pattern, and V_n is the true component of velocity in a direction ϕ_n to the direction of V. V_n' is the apparent velocity associated with a direction at angle ϕ_n to the direction of V; it is equal to x_n/t_n, where t_n is the mean time shift between records obtained at receiving points spaced a distance x_n apart. Thus, if in a practical experiment we determine V_1' and V_2' from pairs of receiving points in directions 1 and 2, V would be given by the construction illustrated in figure 6 (*a*). The main points to be noted are, firstly, that the observed time shifts do not give the true components of V directly, and secondly, that although figure 6 (*a*) illustrates the more usual case in which the axial directions are at right angles, a similar construction can also be applied when the directions are at some finite angle other than a right angle.

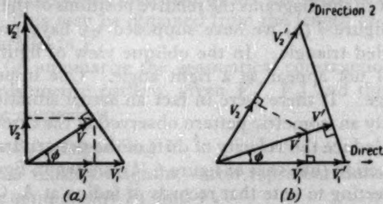


Figure 6. Constructions for obtaining the velocity of an anisometric pattern.

An isometric pattern which is slowly changing in detail as it moves was considered by Briggs, Phillips and Shinn (1950) in relation to axes (Ox and Oy) at right angles; we may, however, generalize their relations in the following manner, so that they may later be applied to pairs of axes at angles other than a right angle.

$$V_n = V \cos \phi_n \qquad \text{......(C)}$$

$$V' = V_n' \cos \phi_n \qquad \text{......(D)}$$

$$VV' = (V_c')^2 \qquad \text{......(E)}$$

Here, in addition to symbols already defined above, we have introduced two extra parameters, V' and V_c'. V' is the apparent velocity deduced from time shifts between records from a pair of receiving points in line with the velocity V of the pattern. An important result of Briggs, Phillips and Shinn is that V' is in general greater than V, and related to it by equation (E). This involves the parameter V_c', discussed in § 2.2, being equal to the ratio of the mean rates of change of amplitude with time at a fixed point, and with distance over the ground at a fixed time.

For our present purpose we note that if we attempt to deduce V by the construction given in figure 6 (a) we arrive at a unique velocity, whatever the directions containing the points of observation, but that the velocity obtained, although it has the direction of V, has a magnitude V'.

Our final summary of the results for an isometric pattern can now take the form of the diagram of velocities shown in figure 6 (b). The axial directions, it will be noted, are not now shown at right angles.

(c) *Relation between an anisometric pattern and the associated isometric pattern.*

We now consider an anisometric pattern such as the one whose details are represented by the contours of amplitude shown in figure 7 (a). We suppose that such a pattern corresponds to a simple stretching in one direction (by a factor r) of an isometric pattern. If this is the case, then it is possible to view figure 7 (a) obliquely in such a way that we appear to see an isometric pattern, as shown in figure 8 (a); this is the pattern that would yield the anisometric pattern if stretched uniformly, by a factor r, in a vertical direction in the diagram.

We can also show in the diagrams the relative positions of the observing points. In the example of figure 7 (a) we have supposed we have points A, O and B forming a right angled triangle. In the oblique view of figure 8 (a), however, the angle AOB will not appear as a right angle. One important point may be noted at this stage. If there were in fact an actual situation corresponding to figure 8 (a), namely an isometric pattern observed at the three points A, O and B, we could readily deduce the velocity of drift of the pattern from observed time shifts by the construction (like that of figure 6 (b)) shown in figure 8 (b). At the same time it is interesting to note that records of fading at A, O and B in such a situation would be identical in form to corresponding records for A, O and B in figure 7 (a) in which the pattern is anisometric.

If we start to draw a similar construction for the case of figure 7 (a) we should first mark off V_1' and V_2' on axes parallel to OA and OB as shown in figure 7 (b). Since, however, amplitude records, and hence time shifts, are identical for figures 7 and 8, the ratio of the velocities V_1' (and V_2') in the two figures is similar to that of the base lines OA (and OB); from this it follows that velocity diagrams, as so far considered, are related (like the other diagrams) by a simple vertical expansion by a factor r.

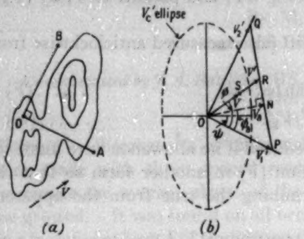


Figure 7. (*a*) A moving anisometric pattern. (*b*) Construction for obtaining the velocity of an anisometric pattern.

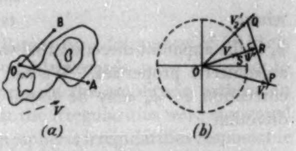


Figure 8. Figure 7 transformed to give an isometric pattern.

Similarly, a vector V in the diagram of figure 7 representing the velocity of drift of the corresponding pattern will have a counterpart in figure 8 which represents the velocity of the isometric pattern. This (being displacement per unit time) will be related to the first vector by a simple vertical expansion like the other vectors in the diagram. A vector V' parallel to V and defined, as given above, in terms of the time shift between records obtained at points in line with V, will behave in the same way. Finally, the ellipse of axial ratio r which represents V_c' is shown in figure 7 (*b*), and appears in the oblique view of figure 8 (*b*) as a circle.

The vector V' is perpendicular to PQ joining the end-points of V_1' and V_2' in figure 8, but the corresponding lines in figure 7 have slopes increased by a factor r; as a result, in figure 7 the vector V' has a slope r^2 times that of the perpendicular ON to the line PQ. In other words, V' and PQ have the directions of conjugate diameters of the ellipse.

The direction of V' gives the direction of V, and we may deduce the magnitude of V from equation (E) above provided we use the value of V_c' appropriate to the direction of V. This may be obtained from the length of the semi-diameter of the V_c' ellipse parallel to V.

We can now (*a*) summarize the geometrical construction needed to find the velocity of an anisometric pattern, given V_1', V_2' and the V_c' ellipse, and (*b*) give the algebraic equivalent.

Geometrical method for determining V.

(i) Draw, in a velocity diagram, vector representations, OP and OQ, of the apparent components V_1' and V_2'.

(ii) The vector representation, OR, of V' is such that R lies on PQ, and OR and PQ are parallel to conjugate diameters of the V_c' ellipse.

(iii) If the V_c' ellipse is drawn centred on O, and OR meets it at T, then OT represents the magnitude of V_c' appropriate to the direction OR.

(iv) The vector representation, OS, of V is such that S lies on OR and $\mathrm{OS}\,.\,\mathrm{OR} = \mathrm{OT}^2$.

Formulae for direction and magnitude of V.

We take direction OP (parallel to OA) as the standard axis, and denote the direction of the minor axis of the ellipse by the angle ψ measured anticlockwise from OA. We denote the values of V_c' along OP, and parallel to V, by $(V_c')_1$ and $(V_c')_v$ respectively.

(i) The direction ϕ of the velocity of drift (also measured anticlockwise from OA) is given by

$$\tan(\phi-\psi)=r^2\tan(\phi_a-\psi) \qquad \text{......(F)}$$

where

$$\tan\phi_a=V_1'/V_2'. \qquad \text{......(G)}$$

ϕ_a is the apparent direction that would be deduced if no allowance were made for anisometric properties. We can put relation (F) in another form so that the correction $\phi-\phi_a$ may be found for determining the true from the apparent direction:

$$\tan(\phi-\phi_a)=\frac{(r^2-1)\tan(\phi_a-\psi)}{1+r^2\tan^2(\phi_a-\psi)} \qquad \text{......(H)}$$

(ii) The magnitude of V', and hence of V, may be deduced from the following relationships which follow from the construction.

$$1/V_a^2=1/V_1'^2+1/V_2'^2 \qquad \text{......(J)}$$

$$V_a=V'\cos(\phi-\phi_a) \qquad \text{......(K)}$$

$$(V_c')_v^2=\frac{1+(r^2-1)\cos^2\psi}{1+(r^2-1)\cos^2(\phi-\psi)}(V_c')_1^2 \qquad \text{......(L)}$$

$$V=(V_c')_v^2/V'. \qquad \text{......(M)}$$

ACKNOWLEDGMENTS

One of us (M. S.) is indebted to the Cavendish Laboratory and St. Catharine's College, Cambridge, for the award of maintenance grants.

REFERENCES

BRIGGS, B. H., PHILLIPS, G. J., and SHINN, D. H., 1950, *Proc. Phys. Soc.* B, **63**, 106.

FÜRTH, R., and MACDONALD, D. K. C., 1947, *Proc. Phys. Soc.*, **59**, 388.

KRAUTKRÄMER, J., 1950, *Arch. Elektr. Übertr.* **4**, 133.

MITRA, S. N., 1949, *Proc. Instn Elect. Engrs*, Pt. III, **96**, 441.

PHILLIPS, G. J., 1952, *J. Atmos. Terr. Phys.*, **2**, 141.

SALZBERG, C. D., and GREENSTONE, R., 1951, *J. Geophys. Res.*, **56**, 521.

SPENCER, M., 1955, *Proc. Phys. Soc.* B, **68**, 493.

The Shape of Irregularities in the Upper Ionosphere

By M. SPENCER
Cavendish Laboratory, Cambridge

Communicated by J. A. Ratcliffe; MS. received 28*th December* 1954 *and in final form* 17*th March* 1955

Abstract. Observations of radio stellar scintillation on a frequency of 38 Mc/s were made with three spaced receivers. Suitable records were analysed to determine the average shape of the irregularities in the amplitude pattern on the ground. It was found on all occasions that the irregularities were elongated by a factor of at least 5 : 1, indicating that the ionospheric irregularities responsible for the amplitude pattern were also very elongated. The directions of stretching were consistent with the hypothesis that the irregularities in the ionosphere were extended along the earth's magnetic lines of force. Further evidence is presented which confirms the hypothesis. It is pointed out that under these conditions it is not possible to determine true velocities of drift from the time shifts between fading curves obtained at spaced receivers, and that the results of previous workers may be seriously in error.

The movement, and the diffusion, of irregularities of the kind here considered are briefly discussed.

§ 1. Introduction

Observations of the fluctuations in intensity of waves from radio stars, analogous to the scintillation of visible stars, have been made by Hewish (1952), Maxwell and Little (1952) and Maxwell and Dagg (1954). They used receiving stations separated by a few kilometres, and compared records of the amplitude of waves from a radio star obtained at the different points. It was found that the variations of phase and amplitude at one point could be ascribed to the steady drift of an irregular wave-pattern over the ground. The irregularities of ionization responsible for the fluctuations were shown to be situated in the terrestrial ionosphere, at a height of about 400 km. From observations of the time shifts between the fading curves, it was deduced that the irregularities drifted with a velocity of the order of 100 to 300 m sec^{-1}.

The work described in the present paper was undertaken to determine whether the ionospheric irregularities responsible for the radio-stellar scintillations were or were not on the average isometric†. In § 2 it is explained what tests can be applied to decide how far a distribution of amplitude on the ground is anisometric. In § 5 these tests are applied to experimental results to show that the irregularities of amplitude on the ground are considerably elongated in a direction roughly North–South, but that the precise direction of the elongation varies with time. An obvious suggestion is that the elongation is observed because the ionospheric irregularities are themselves elongated in the direction of the earth's magnetic lines of force.

---

† The word 'isometric' is used in this paper to denote circular symmetry in a two-dimensional figure and spherical symmetry in a three-dimensional one.

The way in which, under these conditions, the amplitude pattern on the ground would vary as the radio star moved over the heavens is discussed in §4. The apparent directions of drift which would be deduced from the time shifts between fading curves recorded at spaced receivers are also discussed. This theoretical discussion precedes §5 so that the experimental results can be interpreted in the light of the theoretical expectations. It is found that the experiments suggest that the ionospheric irregularities are, in fact, elongated in the direction of the earth's magnetic field.

In §6 the results of previous workers who made observations of radio-stellar scintillations are discussed. In addition, the results given in this paper are compared with the results obtained by Phillips and Spencer (1955), who investigated the lower part of the F region by studying radio waves emitted at the ground and returned after reflection. There is also a short discussion of the possibilities of the diffusion and movement of irregularities in the upper ionosphere.

§2. Tests for Elongation of an Amplitude Pattern

An amplitude pattern in which the irregularities are on the average anisometric can be represented to a first approximation by a 'characteristic ellipse', representing an average amplitude contour or a contour of equal auto-correlation. The constants of the ellipse can be determined from the fading curves recorded at three points.

The instantaneous correlation between two records (that obtained without removing systematic time shifts) gives a measure of the gradient of amplitude over the ground in the direction of the line joining the corresponding receivers. With three receivers the correlation can be evaluated for pairs of points spaced along lines in three directions. The three correlation coefficients can then be converted to radii of an average amplitude contour, as described by Phillips and Spencer (1955). As a first approximation an ellipse can be fitted to the three radii and values obtained for its axial ratio and angle of orientation.

The above analysis can only however be carried out on records which are free from the effects of radio interference. Records which are not good enough for the calculation of cross-correlations can however be used to deduce apparent directions of drift from the time shifts between the different fading curves. We shall therefore examine next the effects of anisometric patterns on these apparent directions.

The calculation of the true velocity of an amplitude pattern from time shifts rests on the assumption that the pattern is statistically isometric, so that the irregularities in it have the same average size in all directions. It is pointed out by Phillips and Spencer (1955) that considerable errors can arise if the pattern is stretched in some preferred direction, so that the assumption is not justified. As the amount of stretching increases, the apparent direction of drift will tend more and more towards the direction of the minor axis of the ellipse. In the extreme case of an infinitely extended pattern, only motion perpendicular to the direction of stretching will be detectable.

If, for example, there existed a systematic stretching of the ground pattern in which the pattern was very elongated in a roughly North–South direction, any North–South components of drift would be almost completely masked.

The East–West components would be preserved, and the apparent drifts would be predominantly East–West. For any assumed variation of the precise direction of elongation it would be possible to predict the approximate variation of the apparent direction of drift, and to test the assumption with experimental results.

§3. THE RELATION OF THE PATTERN ON THE GROUND TO THE IONOSPHERIC IRREGULARITIES

The ionospheric irregularities responsible for radio stellar scintillation on 38 Mc/s introduce variations of phase alone, the absorption being negligible. A pattern of both phase and amplitude is however produced across a plane at some distance from that containing the irregularities. The phase deviations produced in the ionosphere are of the order of one radian, and it has been shown that at normal incidence under these conditions the size of the amplitude irregularities on the ground is the same as that of the irregularities in the ionosphere (Hewish 1951).

When the radio star is away from the zenith, the average shape of the pattern on the ground will be that of the *projection* of the ionospheric irregularities. If they are on the average spherical, their projections on the ground will be on the average elliptical. The pattern on the ground will be anisometric by an amount depending on the angle of incidence of the waves, and the major axis of the characteristic ellipse will vary in orientation with the position of the source. Near upper and lower culmination it will lie roughly North–South, swinging about this direction as the star moves round the sky. For stars of reasonably large declination, the deviation from a North–South line will not be large at any time. There will thus be a tendency for apparent motions deduced from time shifts to be roughly East–West.

The source most frequently observed (23.01) is practically overhead at upper culmination, so that the obliquity effect described will then be small. It is still, however, observed that the apparent motions are mainly East–West. If one is led to suspect that this is still due to anisometric ground patterns, some other mechanism must be postulated.

It is suggested that the irregularities in the ionosphere are elongated along the earth's magnetic lines of force. On this hypothesis it is possible to predict, for a given star and a given time, the orientation of the projection of an irregularity. If the elongation is sufficiently pronounced, one will also be able to predict the direction of the apparent drift deduced from time shifts, since this must be along the minor axis of the characteristic ellipse.

§4. THE PREDICTION OF THE ORIENTATION OF THE CHARACTERISTIC ELLIPSE

We consider the case in which the ionospheric irregularities are infinitely extended along the earth's magnetic lines of force. It is shown in the Appendix that the direction θ (measured eastwards from North) of the minor axis of the characteristic ellipse, when its axial ratio is large, is given by

$$\tan(\theta+\beta) = -\cot(A+\beta) - \cot d \tan\epsilon \operatorname{cosec}(A+\beta) \qquad \ldots\ldots(1)$$

where A is the azimuth of the source, measured eastwards from North, ϵ is the elevation of the source, β the magnetic declination and d the angle of dip. The computation of these angles is also described in the Appendix. When the axial ratio is large, the apparent direction of drift deduced from time shifts can only be θ or $\theta+180°$. A slight change in the true direction of drift, or in the orientation of the characteristic ellipse, can give rise to a reversal of the apparent direction.

The predicted directions θ and $\theta+180°$, calculated from equation (1) for source 23.01 observed at Cambridge, are plotted in figure 1. The directions are plotted as functions of the time interval between the observation and upper culmination of the source. The height of the irregularities was assumed to be 400 km. The curves for 300 and 500 km did not depart from this curve by more than one degree. It will be seen that the directions all lie between 50° and 125°, or 230° and 305°.

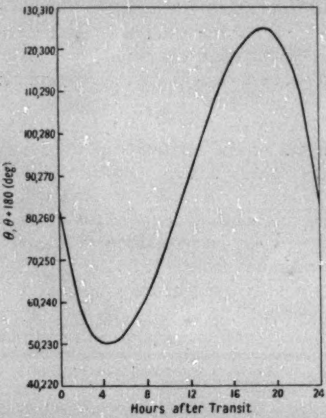


Figure 1. Predicted directions of the minor axis of the characteristic ellipse and the apparent drift deduced from time shifts. Calculated for source 23·01 observed at Cambridge, with irregularities situated at a mean height of 400 km and infinitely extended along the earth's magnetic lines of force.

The case in which the irregularities are not infinitely extended is also treated in the Appendix. It is found that, when the axial ratio of the characteristic ellipse is greater than 10, the errors introduced by assuming an infinite stretching are small.

§ 5. EXPERIMENTAL RESULTS

5.1. *Experimental Arrangement*

The observations to be described were made with three observing stations, situated at the corners of a right-angled triangle. One side was of length 700 m in an East–West direction, and another of length 1200 m in a North–South direction. The apparatus at each station was similar to that employed by Hewish (1952). Observation of source 23.01 (Cassiopeia) was made on a frequency of 38 Mc/s, using a simple interferometer aerial with a phase-switching system (Ryle 1952) to eliminate the effect of background radiation from the galaxy.

5.2. *Determinations of the Characteristic Ellipse*

In this section we report experimental results which were analysed to give a direct measure of how far the amplitude patterns were anisometric. The process described in § 2 was carried out on records which were free from the effects of radio interference. Fading curves obtained at the three receiving points were analysed to determine the instantaneous correlation between the amplitudes. The correlation between one pair of records was often much greater than that between the other two pairs. The effect was almost entirely due to a time shift between one record and the other two. All three records were well correlated when the time shift was removed. In a typical sample of record, the three instantaneous correlation coefficients were 0·41, 0·52 and 0·88. As the greatest correlation was actually that between records from the most widely separated pair of receivers, the result indicated the presence of a very anisometric pattern on this occasion.

The correlation coefficients were in each case converted to radii of a contour of equal correlation, and an ellipse fitted to the three radii. The ellipse then represented approximately an average amplitude contour of the pattern. The ellipse corresponding to the case mentioned above is shown in figure 2. Radii OA, OB and OC are in the directions of the three receiver base lines. The axial ratio of the ellipse was 7 and, as expected from theoretical considerations, the apparent direction of drift deduced from time shifts lay close to the direction of the minor axis of the ellipse.

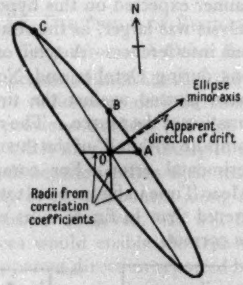

N
C
Ellipse minor axis
B
Apparent direction of drift
A
O
Radii from correlation coefficients

Figure 2. Characteristic ellipse for the example given in the text. OA, OB and OC are radii deduced from correlation coefficients between three pairs of fading records.

The analysis was performed on seven samples of record, taken from five nights' observations in October and November 1954. The records analysed corresponded to experiments each lasting 5 to 10 minutes, during which 5 to 10 maxima of amplitude occurred. The characteristic ellipses were specified by the axial ratio r ($r>1$) and an angle θ giving the direction of the minor axis, measured eastwards from North.† The results are given in table 1, together with the orientations expected if the irregularities were infinitely extended along the earth's magnetic lines of force (equation (1), § 4).

---

† Note that Phillips and Spencer (1955) specified the orientation by an angle measured anticlockwise from a reference line.

It will be seen that the axial ratios were all greater than 5, and the orientations were within 5° of those calculated on the assumption of infinite stretching along the lines of force. The experimental error was of this order.

Table 1

GMT and Date		Time after upper culmination	r	Observed θ	Calculated θ
0025	11.10.54	2h 30m	>20	59°	55°
2357	12.10.54	2h 6m	>20	61°	57°
0017	13.10.54	2h 26m	7	59°	55°
2254	30.10.54	2h 12m	11	61°	56°
0003	30.10.54	3h 23m	>20	52°	51°
2349	8.11.54	3h 44m	>20	52°	51°
0207	22.11.54	6h 56m	5	55°	56°

A large value of r is deduced when the correlation between one pair of records is much greater than that between the other pairs. An equal amount of radio interference at all stations will reduce this great correlation more than it will reduce the small ones, and a smaller value of r will be deduced. The deduced value of r is therefore a lower limit under these conditions.

5.3. *The Variations in Apparent Direction of Drift*

Since the results given in table 1 fitted the hypothesis of highly extended irregularities lying along the earth's magnetic lines of force, all the results available were investigated to see whether the apparent directions of drift deduced from time shifts varied in the manner expected on this hypothesis. The number of records suitable for this analysis was larger, as the results were not so critically dependent on freedom from interference. A total of 43 independent values from 15 nights' observations during October and November 1954 was used. The apparent directions were plotted against the time interval between the observation and upper culmination of the source. The result is shown in figure 3. It will be seen that all the points lie within 8° of the theoretical curve, a departure comparable with the experimental error. For comparison, the results are plotted against Greenwich Mean Time in figure 4. It will be seen that the points are considerably more scattered than in figure 3, as would be expected if the hypothesis here presented is correct.

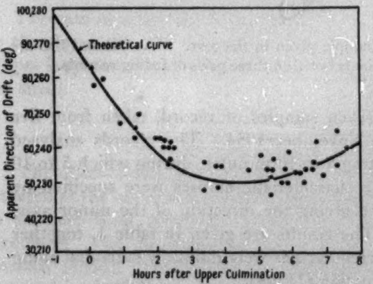


Figure 3. Observed values of the apparent direction of drift deduced from time shifts, plotted against the interval between the observation and upper culmination of source.

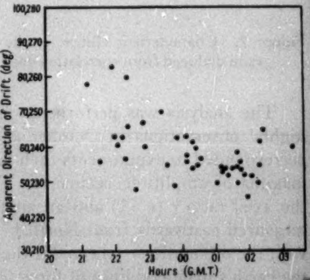


Figure 4. Observed values of the apparent direction of drift deduced from time shifts, plotted against Greenwich Mean Time.

The above results suggest that the irregularities of ionization at heights around 400 km may be considerably extended along the earth's magnetic lines of force on all occasions. Under these conditions it is impossible to deduce the magnitude and direction of the true velocity of an irregularity from time shifts alone. It is, in theory, possible to obtain the true velocity if the constants of the characteristic ellipse are known, but the analysis becomes increasingly difficult to apply as the elongation increases. This is because the constants of the ellipse must be known with very high accuracy. All that can be measured in practice, using time shifts alone, is the component of the true velocity along the minor axis of the ellipse.

5.3. *The Size of the Irregularities*

The records were also analysed to determine the average size of the irregularities in the direction of the apparent motion. The sizes were estimated by dividing the apparent velocity by the number of maxima per unit time, to obtain the average distance between maxima in the pattern. They varied between 2 km and 4 km, and the most frequently occurring value was 3 km. There was no obvious dependence on the position of the source. This indicated that the irregularities were probably in the form of extended spheroids rather than 'curtains' lying in the magnetic meridian. Had the latter been the case, the size would have been markedly dependent on the position of the source. The absence of a marked dependence of this kind was confirmed by the measurements previously made with an East–West pair of receivers by Hewish (private communication).

§ 6. DISCUSSION OF RESULTS

6.1. *Comparison with Other Results*

Since the results here presented indicate that the irregularities of ionization observed with waves from radio stars may be considerably elongated on all occasions, the results of previous workers (Hewish 1952, Maxwell and Little 1952, Maxwell and Dagg 1954) may be seriously in error. If the irregularities are, in fact, elongated on all occasions, then their results give a measure, not of the total velocity, but only of its component along a direction which depends on the position of the source observed. It will be seen that simultaneous observations of two sources would enable the true velocity to be determined. The sudden reversals of apparent direction reported by the workers may, in fact, be caused simply by a slight change in the true direction of drift, or by a steady change in orientation of the characteristic ellipse.

Some experimental results obtained using waves of frequency 2·4 Mc/s reflected at vertical incidence from the ionosphere are given by Phillips and Spencer (1955) in a companion paper. It was found that although on individual occasions the pattern was often anisometric, the effect was not large and did not appear to be systematic. Only three of the records analysed referred to reflection from the F layer, but the largest value obtained for the axial ratio of the characteristic ellipse was only 1·6; a further investigation of this striking difference will be necessary.

6.2. *The Shape of the Irregularities*

We shall make a rough calculation to determine whether anisometric irregularities are to be expected, and by what process they will decay.

We shall first consider the process of diffusion, and shall assume that the effective diffusion coefficients along and perpendicular to the earth's magnetic lines of force are in the ratio of the corresponding mobilities. The ratio is equal to $(\nu^2+\omega^2)/\nu^2$, where ν and ω are the relevant collision frequency and angular gyrofrequency. The diffusion is governed by the collision frequency of ions with neutral particles or other ions, because the electrons cannot diffuse without them and collisions of electrons with ions do not appreciably deflect the ions.

A high acceptable value for the collision frequency of electrons in the F region is 10^3 sec^{-1} (Nicolet 1953). The value may be even less at night. The collision frequency of ions may therefore be taken to be of the order of 10. Putting the gyrofrequency of the ions at 150, the ratio of the diffusion coefficients is then about 200. This large ratio shows that even if an irregularity starts in a spherical form, it will subsequently diffuse out into one which is extended along the earth's magnetic lines of force.

A rough calculation shows that the process of diffusion is probably extremely rapid. An extended irregularity of the size deduced from experimental results will diffuse away in a matter of minutes. The decay by recombination is relatively unimportant; the time constant for this process in the F region is of the order of a day. The rapidity of the diffusion process means that the mechanism responsible for the formation of the irregularities must be in continuous operation while the irregularities are observed.

6.3. *Drifts of the Irregularities*

The result of § 5.3, that the irregularities of ionization appear to be in the form of extended spheriods, makes it likely that the observed drifts are bodily movements of ionization rather than some form of wave motion. One would expect a simple propagated wave to have wave fronts which were extended in two dimensions.

Bodily movement of irregularities of ionization in directions perpendicular to a magnetic field can be produced by an external electric field. In the F region, this is likely to be the only possible mechanism, as mechanical forces will have little effect (Martyn 1953). Clemmow, Johnson and Weekes (1955) considered in detail the case of a cylindrical irregularity lying with its axis along a magnetic field, and found that a steady state solution existed for motion perpendicular to the field.

§ 7. CONCLUSIONS

It has been shown in § 5 that on all occasions when a full analysis was possible, the amplitude patterns produced on the ground by waves from a radio star were found to be highly anisometric. Their orientations were consistent with the hypothesis that the ionospheric irregularities responsible for the patterns were considerably extended along the earth's magnetic lines of force. The hypothesis was confirmed by measurements of apparent drift velocity, deduced from the time shifts between fading curves obtained at spaced receivers. The results of another paper have been quoted to show that true directions of drift cannot be obtained from time shifts when the irregularities are anisometric. The results of previous workers will have to be reconsidered in the light of these facts.

It has been shown in § 6 that a systematic elongation of the irregularities in the upper ionosphere is to be expected. It is also shown that the irregularities may decay by diffusion in a relatively short time, so that a continuous process of formation is required.

ACKNOWLEDGMENTS

I would like to acknowledge the helpful advice and practical assistance of many of my colleagues, workers in both radio astronomy and ionospheric physics. The work described forms part of a programme of work supported by a grant from the Gassiot Committee of the Royal Society. I am indebted to the Cavendish Laboratory and St. Catharine's College, Cambridge, for maintenance grants.

APPENDIX

The Geometrical Problem

Let NWSE (figure 5) be a plane tangential to the earth at the observing point O, and let radiation from the source arrive along lines such as CO, where the angle $COG = \epsilon$ is the elevation of the source, and the angle $GON = A$ is its azimuth, measured eastwards from North. ϵ and A can be derived by spherical trigonometry, which gives the formulae

$$\sin\epsilon = \sin\phi\sin\delta + \cos\phi\cos\delta\cos H$$

and $$\sin A = -\cos\delta\sin H/\cos\epsilon$$

where ϕ is the observer's latitude, δ the declination of the source, and H the hour-angle of the source, measured from upper culmination.

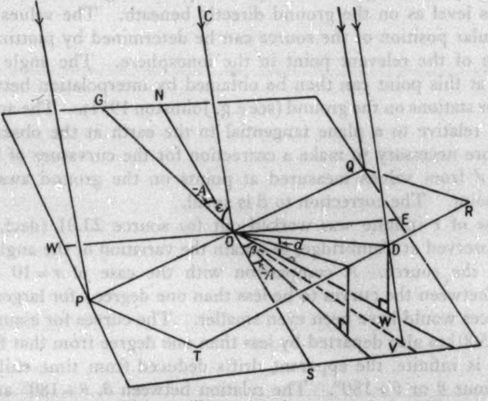


Figure 5. Diagram illustrating the geometrical problem.

Let the line POQ in the plane NWSE be perpendicular to GO, the vertical projection of CO on the plane, and let PQRT be a plane containing a magnetic line of force OD. Let OW be the vertical projection of OD on the plane NWSE, and let the angle $DOW = d$, the angle of dip, and the angle $WOS = \beta$, the magnetic declination. Let OV be the projection of OD on the plane NWSE by rays from the source.

We shall consider first the case of an irregularity which is in the form of a prolate spheroid of axial ratio r, its long axis lying along the line of force OD. Its projection on the plane NWSE will be an ellipse. It can be shown that its minor axis makes an angle θ with ON (measured eastwards) given by

$$\tan 2(\theta+\beta)=\frac{2m(r^2 n\sin d-l\cos d)}{r^2(n^2-m^2\sin^2 d)+(2l^2-1)\cos^2 d-2ln\sin d\cos d}$$

where
$$l=\sin d\sin\epsilon-\cos(A+\beta)\cos d\cos\epsilon,$$
$$m=\cos\epsilon\sin(A+\beta),$$
and
$$n=\cos(A+\beta)\sin d\cos\epsilon+\cos d\sin\epsilon.$$

If r becomes infinite this reduces to

$$\tan(\theta+\beta)=-\cot(A+\beta)-\cot d\tan\epsilon\operatorname{cosec}(A+\beta). \qquad \ldots\ldots(1)$$

In this case the projection lies along OV, and the expression could have been worked out by simple trigonometry.

The angles A and ϵ can be computed for a star of known declination as functions of the time after upper culmination. The time of upper culmination is related to the Right Ascension of the star, and occurs roughly 4 minutes ealier on each successive day.

The angles β and d refer to the region of irregularities through which the radiation from the star passes. We shall assume that the irregularities are located in a region centred at a height of 400 km above the earth. We shall also approximate by assuming that the angle of dip and magnetic declination are the same at this level as on the ground directly beneath. The values of d and β for a particular position of the source can be determined by plotting on a map the position of the relevant point in the ionosphere. The angle of dip and declination at this point can then be obtained by interpolation between values tabulated for stations on the ground (see e.g. Johnston 1951). The angles d and β are defined relative to a plane tangential to the earth at the observing point. It is therefore necessary to make a correction for the curvature of the earth in calculating d from values measured at points on the ground away from the observing point. The correction to β is small.

The case of r infinite was worked out for source 23.01 (decl. 58½°, R.A. 23h 21m) observed at Cambridge, to obtain the variation of the angle θ with the position of the source. A comparison with the case of $r=10$ showed the differences between the curves to be less than one degree; for larger values of r the differences would have been even smaller. The curves for assumed heights of 300 and 500 km also departed by less than one degree from that for 400 km.

When r is infinite, the apparent drifts deduced from time shifts can only have directions θ or $\theta+180°$. The relation between θ, $\theta+180°$ and the time after upper culmination is shown in figure 1.

When r is finite but still greater than unity, the apparent directions will still tend to group around the directions θ and $\theta+180°$. It is possible to calculate the extent of this grouping by the method outlined by Phillips and Spencer (1955). If in a group of results the true directions of drift are uniformly distributed, while the characteristic ellipses have the same orientation and a constant axial ratio of 10, then 96% of the apparent directions will lie within 10° of the direction of the minor axes.

REFERENCES

BRIGGS, B. H. and SPENCER, M., 1954, *Rep. Progr. Phys.*, **17**, 245 (London : Physical Society).

CLEMMOW, P. C., JOHNSON, M. A., and WEEKES, K., 1954, *The Physics of the Ionosphere* (London : Physical Society).

HEWISH, A., 1951, *Proc. Roy. Soc.* A. **209**, 81 ; 1952, *Ibid.*, **214**, 494.

JOHNSTON, H. F., 1951, *J. Geophys. Res.*, **56**, 431.

MARTYN, D. F., 1953, *Phil. Trans. Roy. Soc.* A, **246**, 306.

MAXWELL, A., and DAGG, M., 1954, *Phil. Mag.*, **45**, 551.

MAXWELL, A., and LITTLE, C. G., 1952, *Nature, Lond.*, **169**, 746.

NICOLET, M., 1953, *J. Atmos. Terr. Phys.*, **3**, 200.

PHILLIPS, G. J., and SPENCER, M., 1955, *Proc. Phys. Soc.* B, **68**, 481.

RYLE, M., 1952, *Proc. Roy. Soc.* A, **211**, 351.

The Mechanism of High-Field Electron Emission from Tarnished Metal Surfaces

By T. J. LEWIS
Electrical Engineering Department, Queen Mary College, University of London

MS. received 23rd July 1954 *and in revised form 23rd February* 1955

Abstract. Recent experiments (e.g. Llewellyn Jones and Morgan 1953) have shown that the emission of electrons from smooth cathodes having a thin tarnish film requires the initial deposition of positive ions and that the process then depends on the magnitude of the applied field ($\sim 10^4$ v cm^{-1}). The mechanism of this emission has hitherto been thought to be one of high-field cold emission through the surface barrier. The nature of this barrier is considered in more detail and it is shown that cold emission is improbable. An alternative theory of field dependent thermal emission is proposed which takes into account the existence of patch fields due to variation in the work function over the surface and uses a more realistic value for the degree of field enhancement near the surface. It is shown that, with this mechanism, the predicted emitting areas can be of more probable dimensions.

---

§ 1. INTRODUCTION

It has been shown experimentally (Zuber 1925, Strigel 1939, Tilles 1934, Cobine and Easton 1943, Llewellyn Jones and de la Perrelle 1953, Llewellyn Jones and Morgan 1953) that the time lag to breakdown of small gaps between metal electrodes in various gases depends on the nature of the cathode surface and also on the magnitude of the applied impulse voltage. From these measurements, under appropriate conditions the mean electron current I from the cathode can be deduced. Results obtained by Llewellyn Jones and de la Perrelle (1953) and by Llewellyn Jones and Morgan (1953) for several types of metal cathode surfaces have been expressed in the form

$$I = BE_{\mathrm{m}}^2 \exp(-D/E_{\mathrm{m}}) \qquad \ldots\ldots(1)$$

in which E_{m} is the mean applied field strength ($\sim 10^4$ v cm^{-1}) and B and D are experimentally determined constants.

These authors also found that when an insulating layer in the form of a thin tarnish film is present on the cathode, a copious emission can be obtained provided that positive ions are first deposited on the outside of this surface layer.

The theory suggested by Llewellyn Jones and Morgan (1953) for this case is that cold emission occurs through the potential barrier at the surface. The barrier assumed by these authors is shown in figure 1 (a), in which the underlying metal has a Fermi level of energy μ and work function ϕ_1. Positive ions on the tarnish layer of thickness a produce a lowering of the work function by the amount $\phi_1 - \phi_2$ and further lowering of the barrier beyond this film is produced by the applied field E_{m} enhanced by the factor β due to surface irregularity.

The emission current of electrons through such a barrier was obtained by Stern, Gossling and Fowler (1929) and can be written in the form

$$I = 38{\cdot}5 \times 10^{12} \frac{\mu^{1/2} A \beta^2 E_{\mathrm{m}}{}^2}{(\mu + \phi_1)\phi_1{}^{1/2}} \exp\left(\frac{-6{\cdot}8 \times 10^7 \phi_2{}^{3/2}}{\beta E_{\mathrm{m}}}\right) \frac{\phi_1}{\phi_2}$$

$$\times \exp(-6{\cdot}8 \times 10^7 \psi) \text{ electrons sec}^{-1} \qquad \ldots\ldots(2)$$

where $\psi = \{\phi_1 + \phi_2 + (\phi_1\phi_2)^{1/2}\}a/(\phi_1{}^{1/2} + \phi_2{}^{1/2})$ and A is the effective emitting area of the cathode.

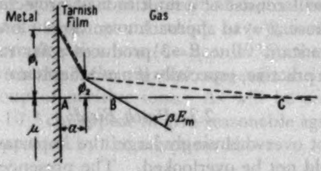


Figure 1 (*a*). Barrier assumed for a cold emission theory.

Interpreting equation (1) in terms of equation (2) and assuming $\beta = 3$, Llewellyn Jones and Morgan have obtained values for ϕ_2 and A of approximately 0·05 ev and 10^{-10} cm^2 respectively for a range of metals in various gases. These values appear unrealistic and for reasons to be discussed it would seem that a cold emission theory may not be justified. An alternative theory which seems more plausible and provides more realistic emission constants will be given.

§ 2. THE SURFACE BARRIER

2.1. *The Tunnel Effect*

In order that the cold field emission mechanism should be adequate it is commonly considered that the width of the barrier AB in figure 1 (*a*) should be of the order of 10 Å, the emission then becoming dependent on the applied field E_{m} when variation in this field produces significant change in the magnitude and width of the barrier. Normally fields greater than 10^6 v cm^{-1} are required (Stern, Gossling and Fowler 1929) to produce the barrier AB. Now the fields employed in the experiments under discussion, even after allowing for considerable enhancement by surface irregularities, are of the order of 10^5 v cm^{-1} and consequently the resultant barrier width AC, instead of AB, figure 1 (*a*), will be too great for tunnelling to take place. The barrier is thus likely to be more closely represented by the dotted curve figure 1 (*a*), than by that pictured by Llewellyn Jones and Morgan (1953) which, having a width AB, is out of proportion.

2.2. *The Field Enhancement Factor*

Various values for β, the field enhancement factor have been used by authors in the past. Llewellyn Jones and Morgan (1953) assume $\beta = 3$. More correctly β is a function of z, the distance from the surface, decreasing to unity at small values of z for most types of surface irregularity within the microscopic range. A useful estimate of $\beta(z)$ can be made by considering a single prolate spheroidal projection on an otherwise smooth plane. Calculations show that while increased sharpness of the projection produces an increased value of $\beta(z)$ at the tip of the projection, at the same time $|d\beta/dz|$ is increased. For instance, a point of height

60 Å and base radius 10 Å gives a value $\beta=24$, $z=0$ and $\beta\simeq 2$, $z=10$ Å, while a hemispherical boss of the same height gives $\beta=3$, $z=0$ and $\beta<2$, $z=20$ Å. These calculations also show that the potential energy change $\alpha(z)$ of an electron in unit applied field is considerably less when $z>30$ Å than that obtained by using the constant value $\beta=3$. In fact, for the hemispherical boss above, the discrepancy becomes appreciable at 15 Å from the surface. While it is true that projections of greater height will extend the range of field enhancement, very much larger projections would seem unlikely for the surfaces under discussion. Furthermore a practical surface will consist of a multitude of projections rather than a single one and this will cause $\beta(z)$ to approach even more closely to unity. Thus the assumption of a constant value $\beta=3$ produces a far more transparent barrier than seems likely in practice, especially if patch fields are taken into account.

2.3. *Patch Fields*

Since βE_m is not overwhelmingly large, the importance of such patch fields (Becker 1935) should not be overlooked. The presence of positive ions on the insulating surface layer produces a lowering of the work function depending on the double-layer strength, so that if the ion deposition is not a uniform one (as indicated below there is good reason to expect this) then the work function will show a similar variation. Since the electron outside the surface must move in a conservative field of force, compensating 'patch' fields will arise so that the total energy barrier surmounted by an emitted electron would be the same whether emitted in a region of high or low work function. These patch fields will be the most important factor in rendering the energy barrier AB (figure 1 (*a*)) inaccurate. Such fields were investigated extensively by Compton and Langmuir (1930) and by Becker (1935). It was shown that the range of the fields was related to the size of the individual patches whilst their magnitudes depended on the strength of the equivalent double layers set up.

It is, of course, impossible to obtain exact expressions for the patch fields for any arbitrary distribution of surface charge but Becker (1935) has derived results for a single circular patch of constant double layer strength on an otherwise uniform surface and also for certain checkerboard arrangements of periodically varying high and low work function patches. Herring and Nichols (1949) have formally taken the problem a stage further by allowing for different periodicities in the two surface directions.

In all these cases the field is found to decrease exponentially with distance from the patch, and at distances large compared with the patch dimensions the patch field is small. For a single circular patch of radius r and double layer strength V_p, Becker has shown that the axial patch field potential energy P_p is given by

$$P_p = eV_p z/(z^2+r^2)^{1/2} \qquad \ldots\ldots(3)$$

provided $z \gg a$, the thickness of the double layer. Although (3) cannot give the true potential, it will be a reasonable approximation when the patches are such that they can be considered individually. This is likely to be so in the experiments under discussion for the following reasons: in the first place the tarnish layer produced as a result of surface oxidation will almost certainly be non-uniform. The rate of oxidation will be affected by local surface conditions and preferred sites for the attachment of oxygen molecules and the subsequent build up of the oxide layer are almost certain to occur (Mott 1947). Secondly if, as seems reasonable, the positive ions deposited on these layers result from previous spark discharges (Llewellyn Jones and Morgan 1953) then the ions will be localized in the neighbourhood of the earlier spark channel. Although Llewellyn Jones (1950) has estimated the spark channel diameter to be approximately 8×10^{-3} cm at the cathode, the diameter of individual positive ion patches will have a high probability of being much smaller than this. There is no evidence that the ions are deposited uniformly from the channel and in any case will, once on the surface, break up and move into equilibrium positions at suitable sites. This break up of the original ion layer will occur as a result of mutual repulsion and clusters might well form around oxygen atoms which have gained electrons from the underlying metal to form negative ions (de Boer 1935), the initial raising of the surface barrier by the presence of the negative oxygen ion being over-compensated by the subsequent group of positive ions. As will be seen later, patch radii of the order of 10^{-5} to 10^{-4} cm appear to give reasonable agreement with experiment and would be supported by the argument above.

2.4. *The Modified Barrier*

Taking into account the above factors and including the image law neglected in figure 1 (*a*), the barrier in the neighbourhood of a positive ion patch will probably be similar to that shown in figure 1 (*b*). With such a barrier cold emission

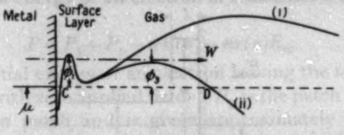


Figure 1(*b*). Barrier in the neighbourhood of a positive ion patch
(i) E_m small. (ii) E_m large ($\sim 10^4$ v cm^{-1}).

(equation 2) would not be expected since the width of the barrier in the presence of the field E_m is CD which is large. If appreciable cold emission is to be obtained E_m must be great enough to overcome the barrier imposed by the patch fields. When this has been accomplished the barrier will approach that shown in figure 1 (*a*). The strength required for such a field can be indicated roughly by the criterion given by Herring and Nichols (1949). E_m must be such that

$$eE_m \gg \frac{\text{difference in work function of the patches}}{\text{patch diameter}}$$

and with a work function difference of the order of 4 ev, E_m must be very much greater than 4×10^4 v cm^{-1} which has not been reached in the experiments.

The barrier of figure 1 (*b*), exhibits two maxima having potentials ϕ_1 and ϕ_3. The maximum ϕ_3 will decrease with increasing E_m and will move in towards the surface at the same time, in a manner which is analogous to the Schottky-type lowering of the work function of a clean surface in a high field, while the maximum ϕ_1 will depend only slightly on E_m. The likely process of electron emission with such a barrier is one in which electrons tunnel through the barrier ϕ_1 and are then emitted over the top of barrier ϕ_3. This would occur for an electron having an energy such as W, figure 1 (*b*), but no emission of an electron from below the Fermi level μ would be possible. Thus the emission process visualized is a thermal one in which the usual emission constant will be reduced by the transmission coefficient of the first barrier ϕ_1. It would be possible to postulate a straightforward Schottky type thermionic emission and to interpret the results of equation (1) without any allowance for a modified surface barrier. Such an interpretation will be found to require values for β and ϕ (the work function of a simple barrier) in the neighbourhood of 5 and 2 ev respectively. These values require further explanation especially as the accepted work functions of the metals used are approximately 4 ev. In this respect the mechanism now suggested provides not only more detail but also acceptable emission constants.

§ 3. HIGH FIELD THERMIONIC EMISSION

In general, for a given barrier, the emission current in electrons cm^{-2} sec^{-1} is given by

$$J = \int_0^\infty N(W) D(W)\, dW \qquad \ldots\ldots(4)$$

in which W is the energy of the electron in a direction normal to the emitting surface, $N(W)$ is the number of electrons in the metal in the energy range W to $W+\delta W$ and $D(W)$ is the transmission coefficient for an electron of energy W, i.e. the probability of such an electron being emitted.

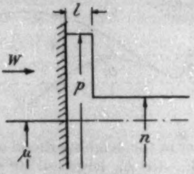


Figure 2. Equivalent rectangular barrier corresponding to figure 1 (*b*).

By representing the barrier of figure 1 (*b*) in the simpler form figure 2, in which the hump ϕ_1 has been replaced by the equivalent rectangular barrier of total height p and width l and the hump ϕ_3 by an equivalent infinitely broad barrier of height n, then, since $D(W)=0$ for $W<n$,

$$J = \frac{4\pi \boldsymbol{mk}T}{\boldsymbol{h}^3} \bar{D}(W) \int_n^\infty \exp\{-(W-\mu)/\boldsymbol{k}T\}\, dW \qquad \ldots\ldots(5)$$

in which $\bar{D}(W)$ is the mean value of $D(W)$ in the range $W \geqslant n$ and the condition $W-\mu \gg \boldsymbol{k}T (T \sim 300^\circ\mathrm{K})$ has been used. For W near to n, i.e. $W=n+\eta$ where η is small, Fowler (1928) has obtained

$$\bar{D}(W) = \frac{8\sqrt{\pi}(\boldsymbol{k}Tn)^{1/2}}{p} \exp[-2\kappa l(p-n)^{1/2}] \qquad \ldots\ldots(6)$$

in which $\kappa^2 = 8\pi^2 \boldsymbol{m}/\boldsymbol{h}^2$.

Since the temperature is low, $N(W)$ will decrease rapidly with increase of W above the Fermi level μ, so that it is justifiable to use equation (6) in equation (5), with the result that

$$J = \frac{32\pi^{3/2} \boldsymbol{mk}^{5/2} T^{5/2}}{\boldsymbol{h}^3} \frac{\sqrt{n}}{p} \exp\left[\frac{\mu-n}{\boldsymbol{k}T} - 2\kappa l(p-n)^{1/2}\right]. \qquad \ldots\ldots(7)$$

While equation (7) applies to the barrier of figure 2, no major modification would be expected for the more correct barrier of figure 1 (*b*). Writing $n=\mu+\phi_3$, $p=\mu+\phi_1$ and $l=a$, then the current density J will be underestimated. This will not be serious, so that substituting numerical values gives

$$J=1{\cdot}55\times10^{26}\frac{(\mu+\phi_1)^{1/2}}{\mu+\phi_3}\exp\left[-38{\cdot}8\phi_3-1{\cdot}02\times10^{8}a(\phi_1-\phi_3)^{1/2}\right]. \quad \ldots\ldots(8)$$

The dependence of J on E_{m} is not explicit in equation (8) but it should be rembered that ϕ_3, occurring at comparatively large distances from the surface, will therefore vary greatly with E_{m}, producing a Schottky-type field dependence which is unlike that for cold emission.

If the suggested emission process is to be tenable, equation (8) should give reasonable agreement with the experimental current-field relationship obtained by Llewellyn Jones and Morgan for particular cathode surfaces. In order to check this a relationship between ϕ_3 and E_{m} is required. With this relationship it will be shown below that satisfactory agreement can be obtained without recourse to exceptional values for $\beta(z)$ or the emitting area A.

§ 4. THE CURRENT–FIELD RELATIONSHIP

4.1. *The Energy Barrier*

The total potential energy of an electron at a distance z from the surface may be written as

$$P=P_{\mathrm{i}}+P_{\mathrm{p}}-eV(z)-e\alpha(z)E_{\mathrm{m}}$$

where P_{i} is the potential energy of an electron leaving the surface when it is free of double layers and with zero applied field. P_{p} is the patch field energy resulting from the positive ion patch and is given approximately by (3). $V(z)$ is the potential function of the patch itself which will vary rapidly through the patch thickness a while the term $e\alpha(z)E_{\mathrm{m}}$ is the energy decrease brought about by the applied field out to the distance z, as in § 2.2.

If a maximum P_{m} of P occurs at $z=z_{\mathrm{m}}$ then

$$\left[\frac{d(P_{\mathrm{i}}+P_{\mathrm{p}})}{dz}\right]_{z=z_{\mathrm{m}}}=e\left(\frac{dV}{dz}\right)_{z=z_{\mathrm{m}}}+eE_{\mathrm{m}}\left(\frac{d\alpha}{dz}\right)_{z=z_{\mathrm{m}}}. \quad \ldots\ldots(9)$$

Also if ΔP_{m} represents the lowering of P_{m} below its value P_{m0} at zero field then

$$\Delta P_{\mathrm{m}}=P_{\mathrm{m0}}-(P_{\mathrm{i}}+P_{\mathrm{p}})_{z_{\mathrm{m}}}+e(V)_{z_{\mathrm{m}}}+eE_{\mathrm{m}}(\alpha)_{z_{\mathrm{m}}}. \quad \ldots\ldots(10)$$

Combining equations (9) and (10) and finally substituting the work function ϕ_{m} corresponding to P_{m} at z_{m} then

$$\frac{d\phi_1}{dE_{\mathrm{m}}}=-e(\alpha)_{z_1}; \qquad \frac{d\phi_3}{dE_{\mathrm{m}}}=-e(\alpha)_{z_3}. \quad \ldots\ldots(11)$$

Confining the attention to the maximum at $z_3(>a)$, and substituting for P_{p} from (3) then equation (9), with $dV/dz=0$, becomes

$$e\frac{3{\cdot}58\times10^{-8}}{z_3^{\,2}}+eV_{\mathrm{P}}\frac{r^2}{(z_3^{\,2}+r^2)^{3/2}}=eE_{\mathrm{m}}\left(\frac{d\alpha}{dz}\right)_{z_3}. \quad \ldots\ldots(12)$$

In equation (12) the normal image law has been substituted for P_{i}. Calculations show that, provided $z_3>40$ Å,

$$E_{\mathrm{m}}=\frac{3{\cdot}58\times10^{-8}}{z_3^{\,2}}+V_{\mathrm{P}}\frac{r^2}{(z_3^{\,2}+r^2)^{3/2}}. \quad \ldots\ldots(13)$$

Finally ϕ_3 can be related to E_m by using equation (10). If ϕ_{30} is the value of ϕ_3 corresponding to P_{m0} at zero applied field, then the reduction of ϕ_3 due to the application of E_m is

$$\phi_{30}-\phi_3=e\frac{3\cdot58\times10^{-8}}{z_3}-eV_p\frac{z_3}{(z_3{}^2+r^2)^{1/2}}+eV_p+eE_m(\alpha)_z, \quad \ldots\ldots(14)$$

in which result, for ϕ_{30}, a value corresponding to the work function of the smooth uncontaminated metal could be taken.

Using equations (13) and (14) a relationship between ϕ_3 and E_m for a given value of r can be obtained provided V_p and $(\alpha)_{z_3}$ are known. From conclusions reached in § 2.2, it is reasonable to assume that $(\alpha)_{z_3}=z_3$ since z_3 will be of the order of 50 Å, while for V_p a value similar to that used by Llewellyn Jones and Morgan (1953), i.e. approximately 4 v will be assumed. Before evaluating the current density equation (8), ϕ_1 and a remain to be discussed.

4.2. *The Values of ϕ_1 and a*

The first maximum ϕ_1 is determined by the magnitude of the double layer V_p, by the image potential, and by the potential change brought about by the field $\beta(z)E_m$. From these potentials by graphical superposition it is possible to show that ϕ_1 lies in the range 3·5–2·5 ev for the representative projections discussed in § 2.2, the maximum occurring at less than 10 Å from the surface.

The table gives calculated values of the current density J from equation (8) for likely values of ϕ_1. It will be seen that J is only significant when ϕ_3 is about 0·5 ev or less and varies very greatly when a is changed from 10 Å to 20 Å. While Llewellyn Jones and Morgan have assumed with some justification that the film thickness can be taken as 20 Å, it is very plausible that this thickness varies over the cathode surface (§ 2.3). Since the surfaces received only a brief exposure to air, the oxidation reaction will not have been completed and the present author is inclined to the view that some patches of thickness less than 20 Å existed.

Using the values $a=10$ Å, $\phi_1=3$ ev calculated values of J and E_m for certain values of r are given in figure 3. The chosen values of r are those most likely to fit the experimental results as indicated below and, at the same time, are reasonable estimates from other points of view.

4.3. *Total Emitting Area*

The emitting area A must be estimated if the experimentally determined total current I is to be compared with the theoretical current density J. If patches exist on the cathode, then the emission will tend to be localized at active sites, the number of such sites increasing with the field E_m. Unfortunately there is no way of assessing the emission from individual patches and one is forced to write $I=J\Sigma A_i=JA$ where J is an average emission density, A_i is the area of the ith patch and A is the total emitting area. Any estimate of A from the measured current I and the theoretically derived values of J must be approximate but nevertheless instructive. If A remains constant, a plot of ln I and ln J against E_m should show the same slope. Figure 4 shows typical results for ln I taken from the paper by Llewellyn Jones and Morgan (1953) and calculated values of ln J from figure 3 for several values of r. In comparing these, it should be realized that the graph for ln I is in fact an average value of several such graphs as those shown for ln J. If however individual curves are compared, say the experimental curve (i) and the calculated current density for a patch radius of 8×10^{-5} cm then the emission area is of the order 10^{-7} cm^2 and increases as E_{m} increases. It is interesting to compare this estimate of the area with that for the spark channel, 5×10^{-5} cm^2 (Llewellyn Jones 1950) over which positive ions are deposited from an individual spark. Since J could be reduced in magnitude considerably by only a slight increase in a (see table), it would be easy to obtain close agreement with the spark channel areas.

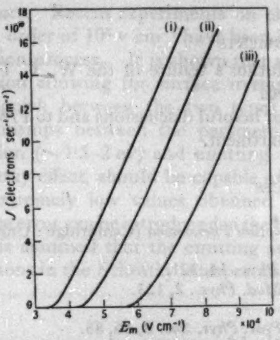


Figure 3. Relationship between current density J and applied field E_{m}.
Patch radius r: (i) 10^{-4} cm
(ii) 8×10^{-5} cm
(iii) 6×10^{-5} cm.

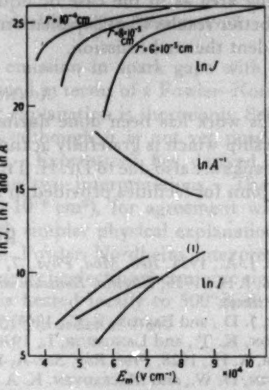


Figure 4. Dependence of ln J and ln I on field E_{m} and derived values of ln A^{-1}.

It should be noticed that the greatest discrepancies in the slopes $d \ln I/dE_{\mathrm{m}}$ and $d \ln J/dE_{\mathrm{m}}$ occur at the lower field strengths. This might be expected, since from equations (8) and (11)

$$\frac{d\ln J}{dE_{\mathrm{m}}} \simeq \left(38{\cdot}8 - \frac{5{\cdot}1 \times 10^7 a}{(\phi_1 - \phi_3)^{1/2}}\right)(\alpha)_{z_0}$$

and will be sensitive to the particular from of patch field chosen in equation (3) when z is large ($\sim$100 Å).

Current density J (electrons cm^{-2} sec^{-1})

ϕ_1 (ev) =		3·5		3·0		2·5	
a (Å) =		10	20	10	20	10	20
ϕ_3 (ev) =	0·10	$5{\cdot}6 \times 10^{15}$	$3{\cdot}9 \times 10^{7}$	$2{\cdot}2 \times 10^{16}$	$6{\cdot}3 \times 10^{8}$	$1{\cdot}0 \times 10^{17}$	$1{\cdot}4 \times 10^{10}$
	0·25	$2{\cdot}5 \times 10^{13}$	$2{\cdot}6 \times 10^{5}$	$1{\cdot}0 \times 10^{14}$	$4{\cdot}7 \times 10^{6}$	$5{\cdot}0 \times 10^{14}$	$1{\cdot}2 \times 10^{8}$
	0·50	$3{\cdot}0 \times 10^{9}$	$6{\cdot}4 \times 10$	$1{\cdot}4 \times 10^{10}$	$1{\cdot}4 \times 10^{2}$	$7{\cdot}5 \times 10^{10}$	$4{\cdot}2 \times 10^{4}$
	1·00	$5{\cdot}1 \times 10$	$5{\cdot}2 \times 10^{-6}$	$2{\cdot}8 \times 10^{2}$	$1{\cdot}6 \times 10^{-4}$	$1{\cdot}8 \times 10^{3}$	$6{\cdot}8 \times 10^{-3}$

§ 5. CONCLUSIONS

Detailed examination of the barrier existing at the surface of a tarnished metal seems to indicate that true cold emission of electrons is improbable with fields of 10^4–10^5 v cm^{-1} because of the likelihood of long-range electron-retarding patch fields. Tunnelling through such a barrier would not be expected to occur especially as calculations indicate that field enhancement is small.

The thermal-type mechanism suggested avoids these difficulties and it predicts more reasonable emitting areas. It is believed that enough has been done to show that, for the range of ϕ_1 and a to be expected, the present theory produces acceptable values for ϕ_3, J, A and $d \ln I/dE_m$ together.

It would appear possible that other instances in which a cold emission process has been assumed and has resulted in unreasonable values of work function and emitting area as in the case of liquids (Goodwin and Macfadyen 1953) might yield better results if an explanation is sought in terms of a Schottky-type field-dependent thermal emission.

ACKNOWLEDGMENTS

This work has been done during the author's tenure of the W. T. Henley Fellowship which is gratefully acknowledged.

Thanks are also due to Dr. H. Tropper for helpful discussions and to Professor W. J. John for facilities provided in the department.

REFERENCES

BECKER, J. A., 1935, *Rev. Mod. Phys.*, **7,** 95.
DE BOER, J. H., 1935, *Electron Emission and Adsorption Phenomena* (Cambridge : University Press).
COBINE, J. D., and EASTON, E. C., 1943, *J. Appl. Phys.*, **14,** 321.
COMPTON, K. T., and LANGMUIR, I., 1930, *Rev. Mod. Phys.*, **2,** 123.
FOWLER, R. H., 1928, *Proc. Roy. Soc.* A, **122,** 36.
GOODWIN, D. W., and MACFADYEN, K. A., 1953, *Proc. Phys. Soc.* B, **66,** 85.
HERRING, C., and NICHOLS, M. A., 1949, *Rev. Mod. Phys.*, **21,** 185.
LLEWELLYN JONES, F., 1950, *Brit. J. Appl. Phys.*, **1,** 60.
LLEWELLYN JONES, F., and MORGAN, C. G., 1953, *Proc. Roy. Soc.* A, **218,** 88.
LLEWELLYN JONES, F., and DE LA PERRELLE, E. T., 1953, *Proc. Roy. Soc.* A, **216,** 267.
MOTT, N. F., 1947, *Trans. Faraday Soc.*, **43,** 429.
STERN, T. E., GOSSLING, B. S., and FOWLER, R. H., 1929, *Proc. Roy. Soc.* A, **124,** 699.
STRIGEL, R., 1939, *Elektrische Stossfestigkeit* (Berlin : Springer).
TILLES, A., 1934, *Phys. Rev.*, **46,** 1015.
ZUBER, K., 1925, *Ann. Phys., Lpz.*, **76,** 231.

Interpretation of Experiments on Electron Emission in Spark Gaps

By M. J. MORANT†
Electrical Engineering Department, Queen Mary College, Mile End Road, London, E. 1

MS. received 7th September 1954 *and in final form* 28*th February* 1955

Abstract. Recent experiments on electron emission in spark gaps with fields of the order of 10^5 v cm^{-1} have been interpreted in terms of a Fowler–Nordheim emission process. It is shown that a better explanation is thermionic Schottky emission allowing for surface irregularities, although it is not yet possible to distinguish between the two types of law. Expressions are derived giving relationships between the parameters in the two interpretations. The work function ($\sim$1·5–2 ev) and emitting area (1–10^{-8} cm^2), for agreement with the Schottky effect, should be capable of a much simpler physical explanation than the extremely low values obtained from a Fowler–Nordheim interpretation. A large error can be introduced in the latter case by neglecting thermionic emission. If it is assumed that the emitting surface is heated locally to 500°c, the work functions in the Schottky interpretation are of the order of 4 ev.

§ 1. Introduction

In a recent series of papers, Llewellyn Jones and de la Perrelle (1951, 1953), Llewellyn Jones and Morgan (1953), and Morgan and Harcombe (1953) have measured the electron emission of various surfaces by observations of the statistical time lag for the electrical breakdown of very small gaps in air and other gases. For particular experimental conditions, the average time lag to breakdown, $\bar{t}$ sec, for a large number of voltage impulses, has been shown to equal the average time interval before the appearance in the gap of a single electron from the cathode (von Laue 1925, Zuber 1925). In this way, very small currents, equal to $1/\bar{t}$ electrons sec^{-1}, can be measured over a small range of field strengths E of about 10^4–10^5 v cm^{-1}. These currents have so far been interpreted as cold (field) emission from the cathode; comparison with theory then gives values of the work function of 0·5–0·02 ev, and emitting areas equal to or less than 10^{-14} cm^2. It is not easy to give a physical significance to these values. It is shown in this paper that the currents are better interpreted as thermionic emission (at room temperature) enhanced by the Schottky effect, and much more reasonable values of the parameters are found. A detailed comparison of the emission equations is given, as similar conclusions may apply to other experiments.

§ 2. The Schottky Effect for an Irregular Surface

The normal equation for the Schottky effect (e.g. Herring and Nichols 1949) assumes a plane surface and uniform applied field. At a distance x from a microscopic projection on an irregular surface, an applied macroscopic field, E v cm^{-1}, will be enhanced by a factor $\beta(x)$, and this modifies the Schottky equation as shown by T. J. Lewis (unpublished). Figure 1 shows the surface image potential barrier of a metal, and the applied field. The peak of the resultant barrier occurs at a distance x_1 from the surface where the force on an electron is zero, i.e. where

---

† Now at The Research Laboratories of the General Electric Co., Wembley, Middlesex.

$$\beta(x_1)=\mathbf{e}/4x_1^2. \qquad \ldots\ldots(1)$$

The image potential at x_1 is $\mathbf{e}/4x_1$, and the potential due to the applied field is

$$\int_0^{x_1}\beta(x_1)E\,dx=\bar{\beta}(x_1)Ex_1. \qquad \ldots\ldots(2)$$

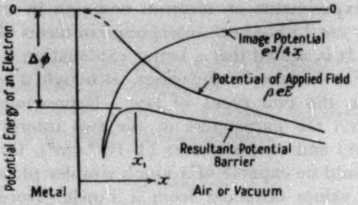


Figure 1. Surface potential barrier of a metal with an enhanced applied field.

Thus the lowering $\Delta\phi_S$ of the work function ϕ_S of the metal, due to the applied field, is

$$\Delta\phi_S=\frac{\mathbf{e}^2}{4x_1}+\bar{\beta}(x_1)\mathbf{e}Ex_1. \qquad \ldots\ldots(3)$$

Substituting for x_1 from (1) gives

$$\begin{aligned}\Delta\phi_S&=\tfrac{1}{2}\mathbf{e}\{1+\bar{\beta}(x_1)/\beta(x_1)\}\{\mathbf{e}\beta(x_1)E\}^{1/2}\\&=\alpha\mathbf{e}(\mathbf{e}E)^{1/2}\end{aligned} \qquad \ldots\ldots(4)$$

where α depends on the surface geometry and is unity for a plane surface ($\beta=\bar{\beta}=1$). The thermionic emission current I over the surface barrier is then

$$I=AT^2\exp\left\{\frac{-\phi_S}{\mathbf{k}T}+\alpha\mathbf{e}\frac{(\mathbf{e}E)^{1/2}}{\mathbf{k}T}\right\}\ \text{A cm}^{-2} \qquad \ldots\ldots(5)$$

where A (theoretically $=120$ A deg^{-2} cm^{-2}) is Richardson's constant, and T is the absolute temperature. A Schottky plot of log I against $E^{1/2}$ has a slope $\alpha\mathbf{e}^{3/2}/\mathbf{k}T$. As α is a function of x_1, given by (1), it is also a function of the applied field strength E. Although not amenable to calculation because of lack of knowledge of the function $\beta(x)$, α is certain to increase with applied field strength as the peak of the potential barrier approaches the surface. Any experimental value of α can be interpreted in several ways in terms of $\beta(x)$ and $\bar{\beta}(x)$. For example, the sets of values in table 1 all correspond to $\alpha=3$ and $E=10^5$ v cm^{-1}. It can be seen that a given value of α is consistent with a high field intensification (and low value of $\bar{\beta}(x_1)/\beta(x_1)$) at a point close to the surface, or a low field intensification at a larger distance. Any conditions between these limits appear to be equally likely.

Table 1

$\beta(x_1)$	1	2	4	6	8
x_1 (Å)	60	42	30	24	21
$\bar{\beta}(x_1)/\beta(x_1)$	5·0	3·2	2·0	1·4	1·1

Experimental Schottky plots are usually made for applied fields of less than 4×10^4 v cm^{-1}. Within this range x_1 is so large that large values of $\beta(x_1)$ and $\bar{\beta}(x_1)/\beta(x_1)$ are unlikely so that α is approximately unity and the field enhancement effect is not normally detected. Even so, as shown in figure 2, Schottky plots may deviate greatly from straight lines because of surface patch fields. For low and moderately high applied fields, the curve has the correct Schottky slope, $e^{3/2}/\boldsymbol{k}T$, and intercepts on the ln I axis of $(\ln AT^2 - \phi_{S\,\mathrm{av}}/\boldsymbol{k}T)$ and $(\ln AT^2 - \phi_{S\,\mathrm{min}}/\boldsymbol{k}T)$ respectively, where $\phi_{S\,\mathrm{av}}$ and $\phi_{S\,\mathrm{min}}$ are the average and minimum work functions of the surface (Herring and Nichols 1949). For applied fields of the order of patch fields, the slope of the Schottky curve is greater than $e^{3/2}/\boldsymbol{k}T$. Although patch fields are usually less than 4×10^4 v cm^{-1}, values of α greater than unity may be partly explained in terms of path effects in fields of up to 10^5 v cm^{-1}.

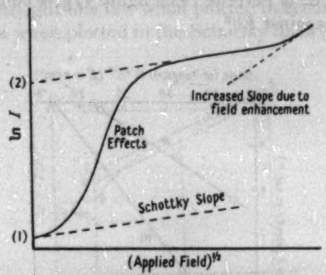


Figure 2. Typical experimental Schottky plot.

§ 3. COMPATIBILITY OF FOWLER–NORDHEIM AND SCHOTTKY PLOTS

The measurements of electron emission referred to in § 1 have hitherto been interpreted in terms of a Fowler–Nordheim field emission law of the form

$$I = BE^2 \exp(-D/E) \qquad \ldots\ldots(6)$$

where B and D are constants. However, since the fields are considerably less than those usually associated with field emission, and the currents are only 10^3–10^5 electron sec^{-1}, thermionic emission over the surface barrier should not be neglected, even at room temperature. With a normal image force barrier, the potential maximum occurs at 15 Å from the surface for the highest field used ($1{\cdot}3 \times 10^5$ v cm^{-1}) and a field intensification, $\beta(15)$, of 12. Thus, tunnelling, even through the top of the barrier, is an unlikely process, and is practically prohibited for electrons below the Fermi level. An explanation in terms of an equation of type (6), derived for a tunnelling process neglecting emission over the barrier, is bound to lead to unsuitable values of the constants B and D, as is shown by the low values of the work function obtained in this interpretation.

Over the small range of applied fields for which the experimental method can be used, plots of log (I/E^2) against $1/E$ give straight lines. However, when a dependent variable changes extremely rapidly with respect to an independent one, several different log plots can usually be found to give straight lines over a small range. In the present case the current can change by as much as 200 times for a 30% change of field, so that agreement with equation (6) is not significant. This can be shown by re-plotting the results in the Schottky form

$$I = AT^2 S_S \exp(\mu + \gamma E^{1/2}) \qquad \ldots\ldots(7)$$

where $\mu = -\phi_S/\boldsymbol{k}T$, $\gamma = \alpha \boldsymbol{e}^{3/2}/\boldsymbol{k}T$ and S_S cm² is the emitting area. As an example, it is shown in figure 3 that results previously plotted as ln (I/E^2) against I/E (Llewellyn Jones and de la Perrelle 1953, figure 4) give equally good straight lines when plotted as ln I against $E^{1/2}$.

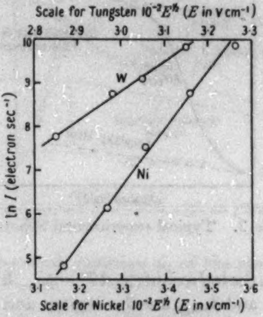


Figure 3. Schottky plot of results of Llewellyn Jones and de la Perrelle (1953, fig. 4).

It can be shown that extremely accurate measurements would be needed to distinguish between (6) and (7) as the correct law for the variation of I with E. Assume that results are available fitting equation (6) exactly so that

$$\ln I = 2\ln E + \ln B - D/E = f(E^{1/2}). \qquad \ldots\ldots(8)$$

We require to find the deviation from a straight line of a Schottky plot (ln I, $E^{1/2}$) of these results. Expanding $f(E^{1/2})$ as a Taylor series in terms of a field E_0,

$$f(E^{1/2}) = f(E_0^{1/2}) + (E^{1/2} - E_0^{1/2}) f'(E_0^{1/2}) + (E^{1/2} - E_0^{1/2}) f''(E_0^{1/2})/2! + \ldots \qquad \ldots\ldots(9)$$

where

$$f'(E_0^{1/2}) = \left(\frac{df}{dE^{1/2}}\right)_{E=E_0} = 4/E_0^{1/2} + 2D/(E_0^{1/2})^3$$

$$f''(E_0^{1/2}) = -4/(E_0^{1/2})^2 + 6D/(E_0^{1/2})^4 \qquad \text{etc.} \qquad \ldots\ldots(10)$$

Therefore

$$\begin{aligned}\ln I = f(E^{1/2}) = f(E_0^{1/2}) &+ (E^{1/2} - E_0^{1/2})(4 + 2D/E_0)/E_0^{1/2} \\ &- (E^{1/2} - E_0^{1/2})^2 (4 + 6D/E_0)/2!E_0 \\ &+ (E^{1/2} - E_0^{1/2})^3 (8 + 24D/E_0)/3!E_0^{3/2} - \ldots \qquad \ldots\ldots(11)\end{aligned}$$

This curve is shown in figure 4. It can be seen that over a small range of $E^{1/2}$ the curvature is very slight. By taking $E_0^{1/2}$ as the mean of the maximum and minimum values of $E^{1/2}$ for a single set of experimental results, it can be shown that the curve, $f(E^{1/2})$, (PQ, figure 4) differs from the straight line

$$\ln I = f(E_0^{1/2}) + (E^{1/2} - E_0^{1/2})(4 + 2D/E_0)/E_0^{1/2} \qquad \ldots\ldots(12)$$

(RS, figure 4) by about 4% of the experimental range of ln I. The actual deviation has been calculated for all the published experimental results quoted previously The largest deviation is 5·9%, the smallest, 0·5%. If the experimental points lay on the curve PQ (figure 4), the best straight line through them would be R′S′ and not RS. This deviates from the curve by about half the above values, i.e. on the average by about 2% of the total range of ln I. The scatter of the experimental points is such that this deviation could not be detected. Thus, experimental results which obey a straight line law when plotted in accordance with (6) would also give a straight line when plotted in the Schottky form (7).

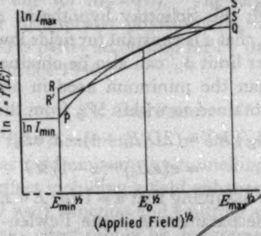


Figure 4. The form of the Fowler–Nordheim equation when plotted as (log $I, E^{1/2}$), and the best straight line approximation.

Equivalence of Parameters of the Two Interpretations

The parameters obtained from the two methods of interpreting the results are, of course, related. From (12) the slope of the Schottky plot in terms of the slope $-D$ of the equivalent Fowler–Nordheim plot is

$$E_0^{-1/2}(4+2D/E_0)=\gamma. \qquad \ldots\ldots(13)$$

Also the intercept of the Schottky line RS (figure 4) on the ln I axis, obtained from (7) and (12), is†

$$\ln(S_S A T^2)+\mu=2\ln E_0+\ln B-3D/E_0-4 \qquad \ldots\ldots(14)$$

where B is the corresponding intercept on the $\ln(I/E^2)$ axis in the Fowler–Nordheim plot. The constants μ, γ, B, and D can be obtained by expressing (6) and (7) in their more usual forms:

Fowler–Nordheim:

$$I=6{\cdot}16\times10^{-6}\,\frac{\zeta^{1/2}}{(\phi_F+\zeta)\phi_F^{1/2}}\,M_F^{\,2}E^2S_F\exp\left(-6{\cdot}8\times10^7\phi_F^{\,3/2}/M_FE\right)\text{amp} \qquad ..(15)$$

Schottky:

$$I=120T^2S_S\exp\left(-\phi_S/\boldsymbol{k}T+\alpha\boldsymbol{e}(\boldsymbol{e}E)^{1/2}/\boldsymbol{k}T\right)\text{ amp} \qquad \ldots\ldots(16)$$

where S_F, S_S cm² and ϕ_F, ϕ_S electron volts are the emitting areas and work functions in the two interpretations of the measurements, and M_F is the field intensification factor for agreement in the Fowler–Nordheim theory. ζ is the kinetic energy of an electron at the Fermi level in the metal, taken as 5 ev. Substitution of the values of the constants in (13) and (14) gives

---

† The use of RS rather than the best straight line R′S′ is much simpler and gives an error of less than 1% in the value of the work function obtained from the intercept.

$$\alpha \boldsymbol{e}(\boldsymbol{e}E_0)^{1/2}/\boldsymbol{k}T = 4 + 13{\cdot}6\times10^7\phi_{\mathrm{F}}^{3/2}/M_{\mathrm{F}}E_0 \qquad \ldots\ldots(17)$$

$$\ln\left\{\frac{6{\cdot}16\times10^{-6}\zeta^{1/2}M_{\mathrm{F}}^2S_{\mathrm{F}}E_0^2}{120T^2S_{\mathrm{S}}(\phi_{\mathrm{F}}+\zeta)\phi_{\mathrm{F}}^{1/2}}\right\} = 4 + \frac{20{\cdot}4\times10^7\phi_{\mathrm{F}}^{3/2}}{E_0M_{\mathrm{F}}} - \frac{\boldsymbol{e}\phi_{\mathrm{S}}}{\boldsymbol{k}T}. \qquad \ldots\ldots(18)$$

Thus the Schottky field intensification factor α is determined uniquely from (17) using values of ϕ_{F}, S_{F}, M_{F}, and E_0 obtained from a Fowler–Nordheim interpretation of a particular set of experimental results. The corresponding work function ϕ_{S} in the Schottky interpretation can be obtained from (18) for various values of the effective emitting area S_{S}. These values of ϕ_{S} are the maximum consistent with the Schottky hypothesis, as it has been assumed, in obtaining the intercept, that α is constant for fields lower than the experimental values. However, a lower limit ϕ_{S}' can also be obtained by assuming that $\alpha = 1$ for applied fields less than the minimum used in any particular experiment. The value of ϕ_{S}' can be obtained to within 5% from the relationship

$$\begin{aligned}(\phi_{\mathrm{S}}-\phi_{\mathrm{S}}')/\boldsymbol{k}T &= (2D/E_0+4) - \boldsymbol{e}(\boldsymbol{e}E_0)^{1/2}/\boldsymbol{k}T\\ &= \boldsymbol{e}(\boldsymbol{e}E_0)^{1/2}(\alpha-1)/\boldsymbol{k}T \qquad \ldots\ldots(19)\end{aligned}$$

which may be derived by assuming that $\alpha = 1$ for $E < E_0$ instead of $E < E_{\mathrm{min}}$.

Typical results for the parameters in the Fowler–Nordheim and Schottky interpretations using values from Llewellyn Jones and de la Perrelle (1953, figure 4) are given in table 2. As will be shown later, the Fowler–Nordheim interpretation is usually incorrect, as the low values of work function required for the correct slope would give a thermionic emission greater than that observed.

Table 2

Material	Fowler–Nordheim Interpretation			Schottky Interpretation			
	M_{F}	ϕ_{F} (ev)	S_{F} (cm²)	α	ϕ_{S} (ev)	ϕ_{S}' (ev)	S_{S} (cm²)
Oxidized nickel	1	0·1	10^{-12}	8·35	2·37	1·46	1
	10	0·47	$1{\cdot}5\times10^{-14}$	8·35	2·14	1·23	10^{-4}
	100	2·3	6×10^{-16}	8·35	1·91	1·00	10^{-8}
Oxidized tungsten	1	0·05	$7{\cdot}3\times10^{-18}$	4·53	1·79	1·39	1
	10	0·23	$1{\cdot}6\times10^{-19}$	4·54	1·56	1·16	10^{-4}
	100	1·10	$4{\cdot}0\times10^{-21}$	4·54	1·33	0·93	10^{-8}

Taking an emitting area, S_{S}, of 1 cm², values of the work functions ϕ_{S} and ϕ_{S}' obtained from a Schottky interpretation of all the published results lie between 1·4 and 2·5 ev. These values are up to 14 times greater than the corresponding work functions ϕ_{F} in the Fowler–Nordheim interpretation. The Schottky values do not depend critically on the assumed emitting area, a reduction of which by 10^8 times changes the apparent work function by only 0·46 ev. Also, they are independent of the field intensification factor α which is uniquely determined and found to have values between 3 and 14 for the wide range of surface finishes to which the published results refer. Values of ϕ_{F} between 0·013 and 0·09 ev for highly polished electrodes of various metals tabulated by Llewellyn Jones and Morgan (1953), indicate that α may be unity and is certainly less than 6 for these very smooth surfaces. Exact values cannot be obtained as the corresponding field strengths are not given by these authors.

In the same way, Kerner and Raether (1954) have recently concluded that direct measurements of electron emission in fields of the order of 10^4 v cm^{-1} can be accurately described as Schottky emission and they obtain similar values for the parameters. These authors assume that the field enhancement does not vary with distance, and so give the slope of the Schottky plot as $\beta \boldsymbol{e}^{3/2}/\boldsymbol{k}T$ instead of $\alpha \boldsymbol{e}^{3/2}/\boldsymbol{k}T$ as derived above.

It is interesting to note that similar conclusions apply to measurements of electron emission into pure liquids derived from the conductivity in fields of up to 2×10^5 v cm^{-1} (Goodwin and Macfadyen 1953). A Fowler–Nordheim interpretation gives extremely low values of work function and emitting area ($\phi_F = 0{\cdot}11$ ev, $S_F \sim 10^{-18}$ cm^2 for $M_F = 10$). The larger values of the Schottky interpretation are easier to interpret physically ($\phi_S \sim 1{\cdot}0$ ev for $S_S = 10^{-14}$ cm^2), but α in this case ($\sim 1{\cdot}3$) is partly accounted for by the dielectric constant of the liquid (LePage and DuBridge 1940).

§ 4. DISCUSSION

It has been shown that measurements of electron emission in spark gaps, previously interpreted as field emission by tunnelling, can also be explained in terms of the Schottky effect including a field enhancement factor which is independent of the values of work function and emitting area. The equivalent work functions for the Schottky effect cannot be determined exactly without knowing the variation of the field intensification factor with distance from the surface. Upper and lower limits of the work function can, however, be obtained, and the difference between them is not great. Although the work functions obtained are lower than the average values for most metals, it is recognized that the intercept of the high-field Schottky plot is determined almost entirely by the patches of lowest work function on the surface (Herring and Nichols 1949). On the Schottky interpretation the effective work function does not vary greatly with the assumed emitting area (1–10^{-12} cm^2) which can be very much greater than the values ($\leqslant 10^{-14}$ cm^2) required for agreement with the Fowler–Nordheim theory.

The normal potential barrier of the Schottky interpretation has been shown to be sufficiently thick for tunnelling to be quite negligible. On the other hand, a Fowler–Nordheim interpretation presupposes a low work function as a barrier must be thin and the applied fields are low. Because of thermionic emission over such a barrier, the Fowler–Nordheim law is definitely incorrect for many of the results quoted. For the example of the nickel and tungsten samples of table 2, the smallest measured currents were $1{\cdot}9 \times 10^{-17}$ and $3{\cdot}6 \times 10^{-16}$ A respectively, and extrapolated to zero field these are $1{\cdot}5 \times 10^{-19}$ and $5{\cdot}0 \times 10^{-18}$ A, and yet the thermionic currents, obtained for the most likely values of work function and emitting area (i.e. for $M_F = 10$) would be $1{\cdot}1 \times 10^{-15}$ and $1{\cdot}8 \times 10^{-16}$ A respectively, even without the Schottky enhancement.

It has been assumed in the above Schottky analysis of the results that the emitting points are at room temperature. However, if the temperature of a single emitting point of area 10^{-5} cm^2 is assumed to be raised to just 500°c, the indicated work functions are of the order of 4 ev and the normal surface barrier of the clean metal is sufficient to explain the results. Local heating of this amount seems possible as the measurement involves 50 discharges per second, and at least 10^{-6} coulomb passed in 1 μsec in a spark channel of diameter of the order of 8×10^{-3} cm (i.e. 2×10^4 A cm^{-2})(Llewellyn Jones and Morgan 1953).

The long time lag before the first breakdown of a series could be explained as the time required for the emission of an electron over a barrier of about 4 ev at room temperature. Subsequently, the emission could be mostly from the same cathode spot, heated slightly by the discharge 2×10^{-2} sec earlier. The discharge could occur at a different part of the surface when an electron was occasionally emitted from outside the cathode spot of the previous discharge.

Variation of the temperature of emitting points would also be consistent with observations of the dependence of the emission on pulse length, obtained at a constant pulse repetition frequency but without spark quenching (Llewellyn Jones and de la Perrelle 1953). These measurements have previously been taken as evidence of a layer of positive ions on smooth surfaces with thin tarnish layers. Such a layer modifies the surface barrier and gives a different interpretation of the constants in the Fowler–Nordheim equation which otherwise gives emitting areas less than molecular dimensions for these surfaces (Llewellyn Jones and Morgan 1953). In view of the reasonable parameters obtained in the Schottky interpretation the use of a more complicated surface barrier than that of the image force does not seem to be justified at present.

ACKNOWLEDGMENTS

The author wishes to express his thanks to colleagues at Queen Mary College for discussion and checking of certain calculations, and to the Department of Scientific and Industrial Research for a grant.

REFERENCES

GOODWIN, D. W., and MACFADYEN, K. A., 1953, *Proc. Phys. Soc.* B, **66,** 85.
HERRING, C., and NICHOLS, M. H., 1949, *Rev. Mod. Phys.*, **21,** 185.
KERNER, K., and RAETHER, H., 1954, *Z. angew. Phys.*, **6,** 212.
VON LAUE, M. , 1925, *Ann. Phys., Lpz.*, **76,** 261.
LEPAGE, W. R., and DUBRIDGE, L. A., 1940, *Phys. Rev.*, **58,** 61.
LLEWELLYN JONES, F., and DE LA PERRELLE, E. T., 1951, *Nature, Lond.*, **168,** 160; 1953, *Proc. Roy. Soc.* A, **216,** 267.
LLEWELLYN JONES, F., and MORGAN, C. G., 1953, *Proc. Phys. Soc.* A, **218,** 88.
MORGAN, C. G., and HARCOMBE, D., 1953, *Proc. Phys. Soc.* B, **66,** 665.
ZUBER, K., 1925, *Ann. Phys., Lpz.*, **76,** 231.

Phase Transitions of the First and Second Orders

By C. HURST
Jesus College, Oxford

MS. received 22*nd November* 1954 *and in amended form* 5*th April* 1955

Abstract. A second-order differential equation in the variables p and T is obtained which is valid for all phase equilibria of the first and second orders. On the basis of this equation, Ehrenfest's treatment of second-order phase changes is presented as a degenerate case of first-order transitions involving discontinuous changes of volume which are very small but non-zero. The ensuing discussion, in which special reference is made to the case of liquid helium, indicates how a fuller understanding of the nature of a second-order transition can be achieved by regarding it as the limiting case of a first-order transition.

---

§ 1. INTRODUCTION

THE fact that the transition from liquid helium I to liquid helium II occurs with no measurable latent heat and with no apparent change of density led Ehrenfest (1933) to investigate the properties of so-called second-order phase equilibria in which both the volume change and the latent heat of transformation at constant pressure and temperature are exactly zero. He showed that, if 1 g of pure substance undergoes a phase transition of this type, then

$$\frac{dp}{dT}=\frac{\Delta C_p}{Tv\Delta\alpha}=\frac{\Delta\alpha}{\Delta k}, \qquad \ldots\ldots(1)$$

where dp/dT refers to the change of pressure with temperature along the second-order transition line, v is the specific volume, and ΔC_p, $\Delta\alpha$ and Δk denote respectively the discontinuities experienced by the values of the specific heat at constant pressure, the volume expansion coefficient, and the isothermal compressibility on crossing the transition line. Objections have been raised by Justi and von Laue (1934), Eucken (1934) and other authors to Ehrenfest's postulate that second-order phase transitions can take place; and, as Rice (1954) has shown, Gorter's model of such transitions, quoted by Keesom (1942), is open to criticism. Since the whole question has evoked considerable discussion, it seems desirable to investigate the connection between first-order and second-order transitions, and to explore the possibility of regarding the latter as limiting cases of the former.

§ 2. GENERAL EQUATION FOR PHASE TRANSITIONS

For a phase equilibrium in which the transformation takes place with the development of finite latent heat L and with a finitely discontinuous volume change Δv, the entropy S satisfies the equation

$$\frac{dS}{dT}=\left(\frac{\partial S}{\partial T}\right)_p+\left(\frac{\partial S}{\partial p}\right)_T\frac{dp}{dT}$$

along the boundary between the phases. In consequence, the discontinuities shown at the boundary by the entropy, the constant-pressure specific heat, and the expansion coefficient α are connected by the relation

$$\frac{d}{dT}(\Delta S)=\frac{d}{dT}\left(\frac{L}{T}\right)=\frac{\Delta C_p}{T}-\frac{dp}{dT}\Delta(v\alpha), \qquad \ldots\ldots(2)$$

where the symbol Δ indicates that the value of the operand immediately to one side of the transition line is to be subtracted from its value at the corresponding point just across the transition line.

Similarly, the equation

$$\frac{dv}{dT}=\left(\frac{\partial v}{\partial T}\right)_p+\left(\frac{\partial v}{\partial p}\right)_T\frac{dp}{dT},$$

which is valid along the phase boundary, implies that

$$\frac{d}{dT}(\Delta v)=\Delta(v\alpha)-\frac{dp}{dT}\Delta(vk). \qquad \ldots\ldots(3)$$

The condition for phase equilibrium is $\Delta G=0$, where G is the Gibbs function. Since $dG=vdp-SdT$, it follows that $dp\Delta v-dT\Delta S=0$ at the boundary. Consequently, if dp/dT exists, the equation $\Delta S=(dp/dT)\Delta v$ holds good for the whole transition line, and can be differentiated along it. On differentiating with respect to T and using equations (2) and (3), we have at once

$$\frac{\Delta C_p}{T}=2\frac{dp}{dT}\Delta(v\alpha)-\left(\frac{dp}{dT}\right)^2\Delta(vk)+\frac{d^2p}{dT^2}\Delta v. \qquad \ldots\ldots(4)$$

In the special case when $L\equiv 0$, and therefore $\Delta v\equiv 0$, the equation (2) becomes

$$\frac{dp}{dT}=\frac{\Delta C_p}{T\Delta(v\alpha)}=\frac{\Delta C_p}{Tv\Delta\alpha},$$

and (3) becomes

$$\frac{dp}{dT}=\frac{\Delta(v\alpha)}{\Delta(vk)}=\frac{\Delta\alpha}{\Delta k}.$$

In this special case, the equation (4) is also consistent with the pair of Ehrenfest equations (1). Thus, it can be regarded as a general equation valid for *all* phase transitions of the first and the second kinds. For equilibria of the third and higher orders it reduces to a trivial identity.

§ 3. PHASE CHANGES WITH NON-ZERO, BUT VERY SMALL, LATENT HEATS

In the general case, equation (4) can be written in the form

$$\left\{\frac{dp}{dT}-\frac{\Delta(v\alpha)}{\Delta(vk)}\right\}^2=\frac{1}{\Delta(vk)}\left\{\frac{d^2p}{dT^2}\Delta v-E\right\}, \qquad \ldots\ldots(5)$$

where

$$E\equiv\frac{\Delta C_p}{T}-\frac{\{\Delta(v\alpha)\}^2}{\Delta(vk)}. \qquad \ldots\ldots(6)$$

By the Ehrenfest equations, $E=0$ and $dp/dT=\Delta(v\alpha)/\Delta(vk)$ when $\Delta v\equiv 0$. Moreover, since in any case of physical interest the volume discontinuity Δv will always be a differentiable function of position along the transition curve, we deduce from (3) that, so long as $\Delta(vk)$ remains finite, dp/dT will tend continuously to the limit $\Delta(v\alpha)/\Delta(vk)$ as Δv tends to zero. For Δv small but non-zero, we therefore write

$$\frac{dp}{dT}=(1+\epsilon)\frac{\Delta(v\alpha)}{\Delta(vk)}. \qquad \ldots\ldots(7)$$

By equation (3),

$$\epsilon = -\frac{1}{\Delta(v\alpha)}\frac{d}{dT}(\Delta v), \qquad \ldots\ldots(8)$$

so that, provided $\Delta(v\alpha)$ and $\Delta(vk)$ remain finite, ϵ can be regarded as a continuous function of Δv which tends to zero as Δv tends to zero. On substituting in (5) from (7) and (8), we now have

$$\frac{d^2p}{dT^2}\Delta v = E + \frac{1}{\Delta(vk)}\left\{\frac{d}{dT}(\Delta v)\right\}^2. \qquad \ldots\ldots(9)$$

It is clear from the above argument that we can now conceive a class of first-order phase equilibria, showing finite discontinuities $\Delta(v\alpha)$ and $\Delta(vk)$, but with non-zero volume changes Δv and non-zero latent heats L so small that ϵ^2 is negligible. It will be convenient to use the term 'small-Δv' to describe an equilibrium of this class. A small-Δv phase transition is therefore one which is governed to a sufficiently close approximation by the equations

$$\frac{dp}{dT} = (1+\epsilon)\frac{\Delta(v\alpha)}{\Delta(vk)} = \frac{L}{T\Delta v}, \qquad \ldots\ldots(10)$$

$$\frac{d^2p}{dT^2} = \frac{E}{\Delta v}. \qquad \ldots\ldots(11)$$

In any specific case of a first-order transition, equation (10) could be used in principle to estimate ϵ from measurements of the physical properties of the phases involved, thus making it possible to judge whether (11) is an acceptable approximation to the exact equation (9). Unfortunately, it is notoriously difficult in practice to measure $\Delta\alpha$ and Δk accurately; as a consequence of this fact, it has not yet been possible, for example, to employ the Ehrenfest equations as a sure test of the second-order nature of the helium transition.

Equation (11) shows that, in the case of a small-Δv transition, the curvature of the boundary between phases remains finite only if E is a quantity of the same order of smallness as Δv. It follows that a substance with the property that the numerical value of E is large for all values of p and T could not exhibit along a transition curve of finite extent a phase equilibrium associated with extremely small values of Δv.

§ 4. SECOND-ORDER PHASE TRANSITIONS

A second-order transition is now to be regarded as the limiting case of a first-order transition with infinitesimally small Δv. By the argument just used, a second-order transition curve of finite extent in the (p, T) plane can be obtained if, and only if, the properties of the two phases happen exactly to satisfy the equation $E_0 = 0$, where

$$E_0 \equiv \frac{\Delta C_p}{T} - \frac{v(\Delta\alpha)^2}{\Delta k}$$

is the expression to which E reduces when $\Delta v \equiv 0$. As soon as the curve runs into a region where $E_0 \neq 0$, it must immediately terminate in a critical λ-point or a triple point. It would indeed be entirely proper to *define* the second-order transition line as the locus of points for which $E_0 = 0$. From this condition, and the first equation (10) with $\epsilon = 0$, we recover the Ehrenfest equations (1).

The circumstance that E_0 must be exactly zero on a second-order transition line has been used by Epstein (1937) as an argument against the existence of second-order transitions, in that a rather unlikely restriction is imposed on the physical properties of the participating phases. In view of the fact that transitions known, or suspected to be, of this type are exceedingly rare in comparison with first-order equilibria, Epstein's argument loses its force once it is realized that no sharp distinction need be drawn between a second-order transition with L, Δv and E all exactly zero, and a first-order transition in which these quantities, though non-zero, have values which are all extremely small.

In the case of the He I–He II transition, the direct experimental evidence in favour of it being either second-order, or so nearly so that the deviation is of no practical consequence, is very good. From the measurements of Keesom and Keesom (1935), it is known that the latent heat developed does not exceed 2×10^{-3} cal g^{-1} at the λ-point. Using the measured slope of the λ-line, it can be deduced from (10) that, if the transition were a small-Δv first-order one, $|\Delta v|$ would have as an upper limit the value 5×10^{-4} cm^3 g^{-1}. Now, a rough estimate, based on experimental values of p and T for points on the λ-line, yields the value $3{\cdot}4 \times 10^7$ dyn cm^{-2} deg^{-2} for $|d^2p/dT^2|$ at the λ-point. Consequently, by equation (11), $|E|$ is, at most, $1{\cdot}7 \times 10^4$ erg g^{-1} deg^{-2}. Since $|\Delta C_p/T|$ is about $3{\cdot}4 \times 10^7$ erg g^{-1} deg^{-2} at the λ-point, the difference there between the values of the two terms forming E is certainly less than 5×10^{-4} times the value of either term taken separately. Thus, although the nature of the case precludes exact experimental confirmation that the transition is true second-order, there can be no doubt that the departure therefrom is small enough to be insignificant in practice.

The arguments already advanced demonstrate that it is legitimate to look upon a second-order transition as the limit attained by the class of small-Δv first-order transitions when Δv tends to zero. In the (p, v) plane, a second-order boundary between phases can therefore be regarded as the result achieved by collapsing laterally the usual ∩-shaped first-order boundary until it becomes one line which is bent back upon itself, as it were, at the critical point. By the time the collapse has been completed the isotherms of the first-order case will have become curves showing a sharp change of slope at the second-order transition line. This correlation between transitions of the first and second orders is quite different from the one adopted by Rice (1954), who draws the diagram of a transition which resembles a first-order transition except that the slope of an isotherm in the two-phase region is not quite zero, and then interprets this diagram as being the representation of two sets of second-order transitions, one set on the left-hand branch of the pseudo first-order boundary and the other on the right-hand branch of this boundary. One obvious defect of this viewpoint is that it seems to require one second-order transition always to be accompanied by another, a prediction which is in conflict with available experimental evidence.

The picture of a second-order transition presented here has the advantage of explaining very simply why liquid helium I and liquid helium II have never been observed to coexist in a volume throughout which there is a uniform temperature and pressure. The normal two-phase region contained within the first-order boundary in the (p, v) diagram has ceased to have an area in the process of collapsing that boundary into a second-order transition line, and consequently there is no longer any possibility of having a phase equilibrium such as is achieved in the first-order case. It is immediately obvious that, in the second-order case, the participating phases can be present simultaneously in a given volume only if one of the variables of state, such as temperature, possesses a gradient in that volume.

References

Ehrenfest, P., 1933, *Leiden Comm.*, Suppl. No. 75 *b*.
Epstein, P. S., 1937, *Textbook of Thermodynamics* (New York : Wiley), p. 131.
Eucken, A., 1934, *Phys. Z.*, **35**, 954.
Justi, E., and von Laue, M., 1934, *Phys. Z.*, **35**, 945.
Keesom, W. H., 1942, *Helium* (Amsterdam : Elsevier), p. 259.
Keesom, W. H., and Keesom, A. P., 1935, *Physica*, **2**, 557.
Rice, O. K., 1954, *Phys. Rev.*, **94**, 1162.

Numerical Ray Tracing in Electron Lenses

By J. C. E. JENNINGS and R. G. PRATT
Birkbeck College, Malet Street, London, W.C.1

MS. received 28th October 1954 *and in final form* 8th *February* 1955

Abstract. Liebmann's method of tracing paraxial rays is critically reviewed, corrections being made to the higher terms of his formulae. A new and very accurate method of ray tracing is described, which is a development of that used by Burfoot; the predictor is one used by Goddard and the corrector is the double integration formula on which the method of Störmer is based. Like Goddard's, this method is applicable to marginal as well as to paraxial rays. The results of numerical tests of various methods are compared. The methods of Gautier and Laudet are discussed. The use of the constancy of the Wronskian determinant as a criterion of accuracy of paraxial ray tracing is considered, examples being given of instances in which it is and is not reliable.

---

§ 1. Introduction

Liebmann (1949) proposed, and he and collaborators (Liebmann and Grad 1951, MacNaughton 1952) have used, a method of paraxial ray tracing and of determining the primary aberrations of axially symmetric magnetic and electrostatic lenses. It is the purpose of the present paper to correct certain errors in Liebmann's paraxial formulae and to compare their accuracy with that of the method of Goddard (1944) and with that of a new method which, like Goddard's, is applicable to the general as well as to the paraxial ray equation.

§ 2. Modification of Liebmann's Method

In Liebmann's method, which is a development of that of Cosslett (1946, pp. 37 and 97), r and r' (here and throughout $f^{(n)} \equiv d^n f/dz^n$) are expanded in Taylor's series :

$$r_{n+1} = r_n + r'_n w + \tfrac{1}{2} r''_n w^2 + \tfrac{1}{6} r'''_n w^3 + \dots \qquad \dots\dots(1)$$

$$r'_{n+1} = r'_n + r''_n w + \tfrac{1}{2} r'''_n w^2 + \tfrac{1}{6} r^{\mathrm{iv}}_n w^3 + \dots \qquad \dots\dots(2)$$

where $w \equiv z_{n+1} - z_n = \text{constant}$.

Both r and r' are computed at every step through the lens; r'' is expressed in terms of r (and of r' if necessary) from the paraxial differential equation, and r''' and r^{iv} are obtained in terms of r and r' by differentiating that equation. Hence equations (1) and (2) may be written in the form:

$$r_{n+1} = r_n Q_1 + r'_n Q_2 \qquad \dots\dots(3)$$

$$r'_{n+1} = r'_n Q_3 + r_n Q_4 \qquad \dots\dots(4)$$

where Q_1, Q_2, Q_3 and Q_4 are functions of w and of the field (and hence of z). In effect Liebmann states (1949, p. 756) that his Q's are correct up to and including w^3, but two errors are committed in their derivation which vitiate this claim. First, although an expression for r^{iv} is derived, it is subsequently neglected; consequently no w^3 terms are given in Q_3 and Q_4, although (apart from Q_3 in a purely magnetic field) such terms are not zero even according to a derivation in which the second error is committed. This second error occurs in the differentiation of r'' and r''', where Liebmann treats the magnetic field H and the electrostatic potential Φ as constants; he states (1949, p. 755) that the resulting error "is cancelled (except for higher order terms)... if the constant field values assigned to the $(n+1)$th interval are the [geometric] mean of the field values at P_n and P_{n+1}". It is true that the use of mean values renders the series for the Q's more rapidly convergent, but although it diminishes it does not eliminate the error arising from treating H and Φ as constants.

Liebmann derives recurrence relations for the primary aberration as well as for the paraxial rays and therefore proceeds from the exact ray equation. For the present purpose it is simpler and sufficient to apply his method to the paraxial equation, which may be written $r''=Ar'+Br$ where A and B are functions of z, but are treated by Liebmann as constants. Such treatment yields

$$r'''=Ar''+Br'=(A^2+B)r'+ABr$$

$$r^{\text{iv}}=(A^2+B)r''+ABr'=(A^3+2AB)r'+(A^2B+B^2)r.$$

Hence

$$Q_1=1+\tfrac{1}{2}w^2B+\tfrac{1}{6}w^3AB$$

$$Q_2=w+\tfrac{1}{2}w^2A+\tfrac{1}{6}w^3(A^2+B)$$

$$Q_3=1+Aw+\tfrac{1}{2}w^2(A^2+B)+\tfrac{1}{6}w^3(A^3+2AB)$$

$$Q_4=wB+\tfrac{1}{2}w^2AB+\tfrac{1}{6}w^3(A^2B+B^2).$$

For a magnetic lens $A=0$, $B=-H^2/4P^2$ where $P\equiv(H\rho)$, the 'specific momentum' of the electrons. For an electrostatic lens $A=-\tfrac{1}{2}\Phi'/\Phi$, $B=-\tfrac{1}{4}\Phi''/\Phi$. Substitution of the appropriate values of A and B in these equations for $Q_1\ldots Q_4$ yields Liebmann's results (1949, equations (18 *a–d*) and (19 *a–d*)) save that, as explained above, the w^3 terms have been omitted from his equations for Q_3 and Q_4, and that there is an obvious typographical error in equation (18 *b*).

We have carried out the differentiations without treating A and B as constants and have thus obtained $Q_1\ldots Q_4$ in terms of H_n and Φ_n. The Q's for a purely magnetic lens are:

$$Q_1=1-\tfrac{1}{8}w^2H_n^2/P^2-\tfrac{1}{12}w^3H_nH_n'/P^2 \qquad \ldots\ldots(5)$$

$$Q_2=w-\tfrac{1}{24}w^3H_n^2/P^2 \qquad \ldots\ldots(6)$$

$$Q_3=1-\tfrac{1}{8}w^2H_n^2/P^2-\tfrac{1}{6}w^3H_nH_n'/P^2 \qquad \ldots\ldots(7)$$

$$Q_4=-\tfrac{1}{4}wH_n^2/P^2-\tfrac{1}{4}w^2H_nH_n'/P^2+\tfrac{1}{12}w^3(H_nH_n''/P^2+\tfrac{1}{8}H_n^4/P^4-H_n'^2/P^2). \qquad \ldots\ldots(8)$$

Following Liebmann (1949, § 11), we may transform these expressions into terms of $H_{n+1/2}$ and its derivatives, in place of H_n etc., where $H_{n+1/2}^2\equiv H_nH_{n+1}$. To effect the transformation we use $H_{n+1}=H_n+H_n'w+\tfrac{1}{2}H_n''w^2+\ldots$ whence

$$H_n^2=H_{n+1/2}^2-H_nH_n'w-\tfrac{1}{2}H_nH_n''w^2\ldots.$$

In these terms the values of the Q's for a magnetic lens are

$$Q_1=1-\tfrac{1}{8}w^2H_{n+1/2}^2/P^2+\tfrac{1}{24}w^3H_{n+1/2}H_{n+1/2}'/P^2 \qquad \ldots\ldots(9)$$

$$Q_2=w-\tfrac{1}{24}w^3H_{n+1/2}^2/P^2 \qquad \ldots\ldots(10)$$

$$Q_3=1-\tfrac{1}{8}w^2H_{n+1/2}^2/P^2-\tfrac{1}{24}w^3H_{n+1/2}H_{n+1/2}'/P^2 \qquad \ldots\ldots(11)$$

$$Q_4=-\tfrac{1}{4}wH_{n+1/2}^2/P^2+\tfrac{1}{12}w^3(\tfrac{1}{2}H_{n+1/2}H_{n+1/2}''/P^2+\tfrac{1}{8}H_{n+1/2}^4/P^4-H_{n+1/2}'^2/P^2). \qquad \ldots\ldots(12)$$

These formulae agree with Liebmann's as far as his go, but it is only in Q_2 that he gives the w^3 term. Comparison between his formulae and our two groups (equations (5)–(8) and (9)–(12)) shows the error of his assumption that the use of $H_{n+1/2}$ justifies the treatment of H as a constant.

For the electrostatic lens a similar derivation yields

$$Q_1=1-\tfrac{1}{8}w^2\psi''+\tfrac{1}{48}w^3\psi''' \qquad \ldots\ldots(13)$$

$$Q_2=w-\tfrac{1}{4}w^2\psi' \qquad \ldots\ldots(14)$$

$$Q_3=1-\tfrac{1}{2}w\psi'+\tfrac{1}{8}w^2(\psi'^2-\psi'')+\tfrac{1}{48}w^3(\psi'''-3\psi'\psi''+3\psi'^3) \qquad \ldots\ldots(15)$$

$$Q_4=-\tfrac{1}{4}w\psi''+\tfrac{1}{16}w^2\psi'\psi''+\tfrac{1}{96}w^3(2\psi^{\mathrm{iv}}+3\psi'^2\psi''-5\psi'\psi''') \qquad \ldots\ldots(16)$$

where for brevity we have written $(\Phi^{(m)}/\Phi)_{n+1/2}\equiv\psi^{(m)}$.

Equations (13)–(16) agree with Liebmann's up to and including the w^2 term, apart from a misprint in the sign of the w term in his equation (19*d*) which is, however, given correctly in equation (7*d*). Liebmann's w^3 terms in both Q_1 and Q_2 are wrong; indeed, that for Q_2 should be zero. He gives no w^3 term in Q_3 or Q_4.

We have followed Liebmann in checking these formulae by applying them to a field in which the paraxial ray equation may be solved analytically. For the magnetic field we have repeated his ray trace (see § 4 below); a simpler check is to calculate the Q's exactly from the analytical solution of the ray equation. From equations (3) and (4) it follows that

$$Q_1=r_{n+1},\ Q_4=r'_{n+1} \text{ for the ray } r_n=1,\ r'_n=0$$

$$Q_2=r_{n+1},\ Q_3=r'_{n+1} \text{ for the ray } r_n=0,\ r'_n=1.$$

The Q's have been calculated to five decimal places for a 'specific momentum' $P=\frac{1}{2}aH_0$ in the field $H=H_0(1+z^2/a^2)^{-1}$ with $w=0{\cdot}2\,a$ (i.e. twice Liebmann's interval) where a is the 'half-width' of the field. The errors in the Q's calculated from equations (9)–(12) and from Liebmann's equations (18*a*–*d*) are displayed in figure 1.

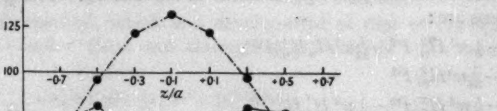


Figure 1. The errors of the Q functions for the field $H=H_0(1+z^2/a^2)^{-1}$, $P=\frac{1}{2}aH_0$; tabular interval $w=0{\cdot}2a$. The error $E(Q)\equiv Q_{\text{approx}}-Q_{\text{analytical}}$. Q_2 and Q_4 are symmetrical about $z/a=-0{\cdot}1$. The formulae of the present authors and of Liebmann are identical for Q_2.

A similar but less extensive check has been made on the formulae for the electrostatic field, by applying them to the field $\Phi = \Phi_0 \exp[(4/\sqrt{3}) \tan^{-1}(z/a)]$ for which Hutter (1945) gave the paraxial solution

$$r = C \operatorname{cosec} \beta \sin(\sqrt{2}\beta + \alpha) \exp[(\beta - \tfrac{1}{2}\pi)/\sqrt{3}]$$

where C and α are arbitrary constants and $\cot \beta = z/a$. Table 1 shows for $z/a = 5{\cdot}0$, $w = 0{\cdot}5a$ and $1{\cdot}0a$ the analytical values of the Q's together with the errors of those obtained from Liebmann's formulae (19 *a–d*) and from our formulae (13)–(16). (Error $\equiv Q_{\text{approx}} - Q_{\text{analytical}}$).

Table 1

$w=0{\cdot}5a$	Q_1	Q_2	Q_3	Q_4
Analytical value	1·000 754	0·494 936	0·980 636	0·002 879
Error of Liebmann's formula	−0·000 030	+0·000 135	−0·000 038	+0·000 018
Error of our formula	−0·000 001	−0·000 021	+0·000 005	−0·000 001
$w=1{\cdot}0a$				
Analytical value	1·002 772	0·981 385	0·965 785	0·005 118
Error of Liebmann's formula	−0·000 142	+0·000 837	−0·000 216	+0·000 109
Error of our formula	+0·000 063	−0·000 290	+0·000 082	−0·000 018

Liebmann's formulae for the magnetic lens are simpler than ours and if paraxial ray tracing alone were in question, it would be preferable to reduce the tabular interval w rather than to use our w^3 terms with the wider interval; even so, our formulae would be valuable for giving an indication of the error in the Q's at any given tabular interval. As shown in § 4, however, the entire procedure is disadvantageous for purely paraxial ray tracing.

If the primary aberration is to be determined by Liebmann's method, our equations (9)–(16) will be preferable to his for the paraxial trace, since the derivatives required in our w^3 terms have in any case to be evaluated for Q_5, the the aberration term (Liebmann's equations (18 *e*) and (19 *e*)).

§ 3. A SIMPLIFIED METHOD OF PRECISE RAY TRACING

It is well known that the general ray equations in an axially symmetric field may be written as a simultaneous pair : $d^2r/ds^2 = \phi_1(r, z)$, $d^2z/ds^2 = \phi_2(r, z)$ (Maloff and Epstein 1938, eqn. (8.17)). Since the problems of solving a single non-linear equation or a simultaneous pair are not vastly different, we shall discuss the single equation with first derivative absent:

$$r'' = \phi(r, z). \qquad \ldots\ldots(17)$$

The paraxial equation, whether there is axial symmetry or not, is always linear and hence may be reduced (if necessary by a change of dependent variable, cf. Picht 1932) to the form

$$r'' + rf(z) = 0. \qquad \ldots\ldots(18)$$

The very accurate central difference method of solving equations of the types (17) and (18) which was proposed by Störmer (1907) is the basis of our own. As described by Hartree (1952, p. 127) and Milne (1953, p. 86), this method is open to the objection that it involves estimation of δ''† one place ahead, of δ^{iv} two places ahead, of δ^{vi} three places ahead (and so on) of the last available value. Although Comrie (1944) argued strongly for this method, many have found it prohibitively tedious unless the tabular interval w is so reduced that the term in δ^{iv} is negligible or small.

---

† We use $\Delta^{(n)}$ to denote the nth central difference of r, $\delta^{(n)}$ to denote that of r''.

Backward difference formulae are not open to this objection but their convergence is extremely poor. Booth (1954) has pointed out that as early as 1883 Bashforth and Adams proposed a combination of backward and mixed difference formulae, for the integration of first order differential equations. They used a purely backward difference formula to predict one step ahead, formed differences from the predicted result and then corrected it with the help of a more convergent mixed formula. Milne (1933) applied the same technique to equation (17), his predictor being

$$r_{n+1} = r_n + r_{n-2} - r_{n-3} + \tfrac{1}{4}w^2(5r''_n + 2r''_{n-1} + 5r''_{n-2}) + \sim(17w^2/240)\delta^{\mathrm{iv}} \qquad \ldots\ldots(19)$$

which was followed by the corrector

$$r_{n+1} = 2r_n - r_{n-1} + \tfrac{1}{12}w^2(r''_{n+1} + 10r''_n + r''_{n-1}) - \sim(w^2/240)\delta^{\mathrm{iv}}. \qquad \ldots\ldots(20)$$

If one eighteenth of the discrepancy between predicted and corrected values is significant, the tabular interval w must be reduced, or a formula of higher order used. Goddard (1944) based a method of ray tracing on equation (19) alone.

Fox and Goodwin (1949) have proposed, and Burfoot (1952) has applied to paraxial electron optics, a method whereby the predictor is used throughout the trace and then followed by the corrector, instead of their being applied alternately at each step. This has the advantage that high order difference corrections may be brought in, enabling w to be relatively large without loss of accuracy. In particular, this procedure gives flexibility in the order of differences used, which often obviates the necessity of changing w throughout a trace. This flexibility, however, depends upon the coefficients of high order terms being known, as they are in equation (24) below.

The general procedure of Fox and Goodwin, when applied to first order equations, makes possible the use of the most convergent of all formulae for single integration between tabular points

$$\int_{-w}^{+w} y\,dx = 2w\left(y_0 + \tfrac{1}{6}\Delta''_0 - \tfrac{1}{180}\Delta^{\mathrm{iv}}_0 + \tfrac{1}{1512}\Delta^{\mathrm{vi}}_0 - \tfrac{23}{226800}\Delta^{\mathrm{viii}}_0 + \tfrac{263}{14968800}\Delta^{\mathrm{x}}_0 - \,..\right). \qquad \ldots\ldots(21)$$

(see Comrie 1947, p. 942). This form is applicable alike to linear and non-linear equations, whereas the correctors of Fox and Goodwin are for linear equations only and are less convergent than (21).

For the linear second order equation (18), equation (20) may be rewritten

$$r_{n+1}(1 + \tfrac{1}{12}w^2 f_{n+1}) = r_n(2 - \tfrac{5}{6}w^2 f_n) - r_{n-1}(1 + \tfrac{1}{12}w^2 f_{n-1}) + D \qquad \ldots\ldots(22)$$

$$D = -\tfrac{1}{240}\Delta^{\mathrm{vi}}_n + \tfrac{13}{15120}\Delta^{\mathrm{viii}}_n - \tfrac{17}{100800}\Delta^{\mathrm{x}}_n + \tfrac{1}{29700}\Delta^{\mathrm{xii}}_n - \,\ldots\ldots \qquad \ldots\ldots(23)$$

Fox and Goodwin and, following them, Burfoot, use as predictor equation (22) with $D = 0$ and as corrector (22) with D given by (23).

We use the less convergent (19) as predictor, difference r'' and use (24) as corrector, re-calculating r'' from the corrected r at each step. The size of the discrepancy between predicted and corrected r'' indicates whether a second application of the corrector is needed and whether δ'' should be revised ; we have rarely found this necessary and have never needed to revise the higher differences. Irregularity of the discrepancy between predicted and corrected values of r reveals mistakes, although small mistakes may be masked if the truncation error of the predictor is oscillatory.

Our corrector is the central difference formula for double integration on which Störmer's method is based:

$$\Delta_n'' = w^2(r_n'' + m_2\delta_n'' + m_4\delta_n^{\mathrm{iv}} + m_6\delta_n^{\mathrm{vi}} + m_8\delta_n^{\mathrm{viii}} + \dots) \qquad \dots\dots(24)$$

where

$$\begin{array}{ll} m_2 = +0{\cdot}08333\ 33333 & m_{12} = -0{\cdot}00000\ 26014 \\ m_4 = -0{\cdot}00416\ 66667 & m_{14} = +0{\cdot}00000\ 05104 \\ m_6 = +0{\cdot}00051\ 25661 & m_{16} = -0{\cdot}00000\ 01036 \\ m_8 = -0{\cdot}00007\ 96407 & m_{18} = +0{\cdot}00000\ 00216 \\ m_{10} = +0{\cdot}00001\ 38977 & m_{20} = -0{\cdot}00000\ 00046. \end{array}$$

We have calculated the coefficients from the exact values given by Oppolzer (1880, p. 545) and reproduce them because of the inacessibility of that work. It is a trivial transformation to convert (24) to a formula for r_n in terms of the double sum of r_n'', r_n'' itself and its central differences (cf. Comrie 1947, p. 809); this is sometimes preferable as a precaution against building-up errors.

Two modifications to the procedure just described are sometimes needed. First, if, as in the lens field of a solenoid whose length considerably exceeds its diameter, there are two ranges of z where $f(z)$ in equation (18) varies rapidly, it pays to stop the predicted trace in the middle of the lens and to correct it fully before applying the predictor to the second half. Secondly, in a diverging lens (i.e. $f(z)<0$) a converging solution (i.e. one of the order of e^{-z} in a region where $f(z)\sim$const.) may become contaminated and swamped by the e^z solution; in so far as this contamination arises from truncation error of the predictor, it ceases to be advantageous to carry the predicted trace many steps ahead. This problem has been discussed by Todd (1950). An example of such a diverging electron optical system is a magnetic prism having a plane of anti-symmetry (Khurgin 1939, Sturrock 1950).

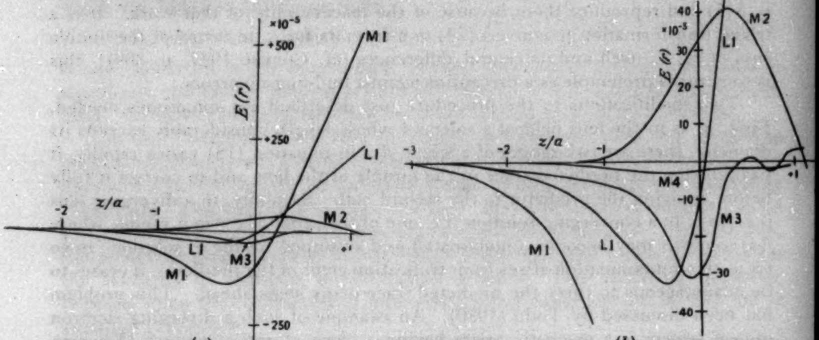
The advantages of our method over that of Burfoot are (i) it applies to non-linear as well as to linear equations; in this connection, both our predictor and our corrector are more convergent than the formulae (4.13) mentioned by Fox and Goodwin; (ii) the predictor (19), although less convergent and using four instead of two known points, needs the tabulation of only one function (f), whilst Burfoot's (our equation (22)) needs two ($1+\frac{1}{12}w^2f$ and $2-\frac{5}{6}w^2f$); (iii) the predictor (19) involves no division; it is especially simple if $w=2\times 10^n (n=0, \pm 1, \pm 2 \dots)$; (iv) our corrector (24) is about twice as convergent as the corrector (23); (v) coefficients in equation (23) have been published only as far as that corresponding to (but exceeding) m_{10} whilst for equation (24) we have m_{20}. Working to six figures in r, which is not excessive if aberration is to be found from the difference between marginal and paraxial traces, we found it necessary to use coefficients as far as m_{16} in part of the trace to which table 2 (viii) refers.

It is, of course, true that a predictor–corrector sequence involves duplication of labour, but it may well be preferable to a single trace with half the tabular interval, as the additional work involved in halving the interval is more than a mere doubling, for the field must be calculated at the intermediate points. Even if this can be done by subtabulation, the labour is not negligible; but it is not obvious that subtabulation is always feasible. In our method, as in Burfoot's, the second trace is independent of the first, which affords a check on mistakes, not available in the labour-saving 'η-process' of Clenshaw and Olver (1951).

§ 4. COMPARISON OF METHODS

A direct check on the accuracy of marginal ray tracing is scarcely feasible for lack of a suitable field having an analytical solution to the ray equation. Comparison has therefore been made on the paraxial equation.

Liebmann checked his magnetic lens formula by tracing a ray of momentum $P = \frac{1}{2}aH_0$ in a field $H = H_0(1+z^2/a^2)^{-1}$, with initial conditions $r=1$, $r'=0$ at $z/a=-3{\cdot}1$; he used an interval $w=0{\cdot}1a$ and compared the z-intercept and the slope r' at that point with the values calculated from Glaser's (1941) solution. It is more instructive, however, to make the comparison throughout the lens, as figure 2 shows.



(*a*) (*b*)

Figures 2 *a* and *b*. The errors of the ordinate of the ray $r_{-3\cdot1}=1$, $r'_{-3\cdot1}=0$, $P=\frac{1}{2}aH_0$ in the field $H=H_0(1+z^2/a^2)^{-1}$. $E(r)\equiv r_{\text{approx}}-r_{\text{analytical}}$. L1 denotes Liebmann's (1949) trace, with $w=0{\cdot}1a$. M1–4 denote methods 1–4 for which $w=0{\cdot}2a$.

We have traced the same ray with $w=0{\cdot}2a$ by means of:

Method 1: Liebmann's formula (1949, equations (18*a*–*d*)),
Method 2: our modification of Liebmann's formula (equations (9)–(12) above),
Method 3: Goddard's formula (equation (19) above),
Method 4: our predictor–corrector method, method 3 providing the first approximation. To the accuracy employed no further change would have resulted from a second application of equation (24).

The error in r is shown in figures 2 (*a*) and 2 (*b*), which also show the error in the trace of the same ray by Liebmann (1949, pp. 768–9) with $w=0{\cdot}1a$. Two figures are needed because of the great range of the magnitudes of the errors of the various methods. Throughout we took r to five decimal places, save that in method 4 we carried a non-significant sixth figure in Δ'' and r, purely as a protection against accumulation of rounding errors. Whilst, in methods 1 and 2, H^2 was calculated, following Liebmann, to five decimal places, in methods 3 and 4 only four places were used; in spite of this the accuracy of the trace by method 4 is far superior to that by the other methods. Figure 3 exhibits the errors in r' for methods 1 and 2, and for Liebmann's own trace with $w=0{\cdot}1a$.

In figure 2 (*b*) the errors in the traces by methods 2 and 3 are about equal in magnitude. This is fortuitous, however; the leading error terms in Δ' are

$\sim w^4 r^{\rm iv}$ and $\sim w^6 r^{\rm vi}$ respectively. In fields which vary slowly enough, therefore, Goddard's method must give greater accuracy than equations (9)–(12).

It is pertinent to add that, although such high accuracy as has been achieved with method 4 is wanted only rarely, the method should show advantage in labour and accuracy also when fewer decimal places are required, so that a still wider interval could be used.

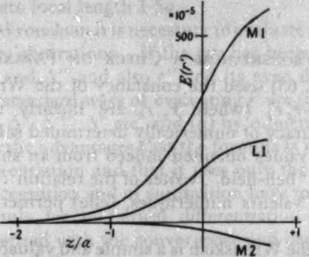


Figure 3. The error of the slope of the ray $r_{-3\cdot1}=1, r'_{-3\cdot1}=0$, $P=\frac{1}{2}aH_0$ in the field $H=H_0(1+z^2/a^2)^{-1}$. $E(r')\equiv r'_{\text{approx}}-r'_{\text{analytical}}$. L1 denotes Liebmann's (1949) trace with $w=0\cdot1a$. For the other traces $w=0\cdot2a$.

Two further methods of numerical ray tracing have recently been described by Gautier (1953) and Laudet (1953). Gautier uses for the paraxial ray (he also considers primary aberration) the formula $r_{n+1}=2r_n-r_{n-1}+w^2r''_n$ in which the leading error term is $\frac{1}{12}w^4 r^{\rm iv}$. This is justly described by Milne (1949, p. 140) as a ' crude but rapid method '. Its error is of the same order as that of equations (9)–(12), whilst it requires only two preceding points compared with four for equation (19). Gautier traces what purports to be the same ray as that of Liebmann and ourselves, but in fact he makes the boundary conditions $r_{-3\cdot1}=r_{-3\cdot0}=0\cdot1$ so that comparison cannot be made.

Laudet proposes several formulae for paraxial tracing, of which the most accurate, according to a trial in Glaser's ' bell-field ' with $2P=aH_0/\sqrt{3}$, is equation (22) above with $D=0$.

He makes traces at intervals w and $2w$ (the size of w is not stated) and uses the deferred approach to the limit (Richardson 1927, Duncan 1948) to get a better approximation, whose error is only two or three times that of method 4. He takes the index (Duncan 1948) of his formula to be 3, which would correspond to an error in Δ' of order w^4, whilst the error is known to be of order w^6. It is hard to resist the conclusion that the good final result is fortuitous, in much the same way as the approximate equality of the errors in methods 2 and 3, illustrated in figure 2 (*b*).

We have also found difficulty in understanding the choice of index adopted by Liebmann (1949, § 9) when using the deferred approach; he assigns different indices to r and r', which seems to us unjustified since each quantity is calculated from the preceding values of both, at every step. Figures 2(*a*) and 3 show that the ratio of the errors of the traces L1 and M1 (Liebmann's method with $w=0\cdot1a$ and $0\cdot2a$) varies greatly with z, both for r and r'; in other words, the deferred approach is not reliable in this case. We believe that for the deferred approach to be valid the formulae must be such that the indices for r and r' are the same (e.g. our equations (9)–(12)). However, the method of the deferred approach appears to us to be at best uncertain, for although in its favour it may be said that the independence of the traces affords a check on mistakes, yet for the extrapolation to be valid the truncation error from all but the leading term must be negligible at the wider interval, so that the smaller interval must involve very many steps. Further, there is no simple criterion by which the final accuracy can be estimated.

§ 5. THE WRONSKIAN AS A CHECK ON PARAXIAL ACCURACY

Dietrich (1951, p. 63) used the constancy of the Wronskian determinant of equation (18), $r_a r_b' - r_a' r_b$ (where r_a, r_b are linearly independent solutions), as a check on the accuracy of numerically determined solutions. Laudet (1953), regarding numerical values obtained indeed from an analytical formula such as that of Glaser for the ' bell-field ', writes of the relation : Wronskian = constant, that " appliquée aux valeurs numériques, (elle) permet de mettre en évidence toute erreur de calcul ".

The constancy of the Wronskian is a simple and valuable check on the freedom of a numerical solution from the cumulative effects of rounding errors and from mistakes; it may also be a check against truncation errors, but its sensitivity in this respect depends on the formula used for integrating the differential equation. For the constancy of the Wronskian is not a sufficient, although it is a necessary, condition that a given function satisfies equation (18); indeed, at least for a field symmetrical about $z=0$, it may be proved that the Wronskian of a solution found by Liebmann's method is identically constant, whatever the tabular interval and, hence, whatever the truncation error. The proof depends on the symmetry of the Q's about $z=-\frac{1}{2}w$ and on the equality of Q_1 to Q_3.

Although truncation error is not detected by the Wronskian check in this case, it is detected when some other numerical methods are employed for solving the paraxial equation (18); in the absence of a general theory of this useful property of the Wronskian, help may be derived from a tabulation of some results of this check; these are given in table 2.

Sections (i) to (vi) of table 2 refer to a ray traced through Glaser's ' bell-field ', as described in § 4; the error of the trace by the various methods is shown in figure 2. Comparison between table 2 and figure 2 confirms that the Wronskian test is useless as a check on truncation error when applied to method 1 and shows

Table 2. Values of Wronskian Determinant

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	z	(vii)	(viii)
n	†L1	M1	M2	Glaser	M3	M4	(cm)	M3	M4
0·5	0·73827	0·73485	0·74075	0·740077	0·73928	0·74004	0	3·57349	3·57491
1·5	7	5	5	77	0·74095	4	10		1
2·5	6	5	5	80	0·74005	6	26	3·57519	0
3·5	6	5	4	78	0·73873	6			
4·5	7	5	3	78					
5·5	7	5	3	76			z/a	(ix)	(x)
6·5	6	5	4				0·0	0·63736	0·63748
7·5	7						0·5	0·63750	8
8·5	6						1·0	0·63719	8

† In (i) for n read $n-\frac{1}{2}$, e.g. the first entry is at $n=0$. Throughout the table $n=z/w$.

that it is extremely insensitive for method 2, but a useful guide for methods 3 and 4. Sections (vii) and (viii) refer to traces (with $w = 2$ cm) of one ray (for which the focal length was approximately 28 cm) through the lens field of a solenoid of length 36 cm and inner and outer radii 3·810 cm and 6·386 cm respectively; calculation of the field and tracing of the ray were performed by Mr. D. K. Das Gupta and one of us (J.C.E.J.). Sections (ix) and (x) refer to traces (with $w = 0{\cdot}2a$) of one ray through the lens field of a solenoid having a single layer of radius a, length $2a$ and approximate focal length $1{\cdot}5a$.

In order to form the Wronskian it is necessary to evaluate r'; this is also needed when evaluating primary aberrations. If the ray has been traced by method 4, r and its differences Δ' and Δ'' and also r'' and its even differences δ'', δ^{iv} etc. will be available. The standard ways of evaluating r' are (i) by numerical differentiation of r, viz. from Δ', Δ''', Δ^{v} etc. and (ii) by numerical integration of r''. The second method has the advantages that the formula is more convergent than that for numerical differentiation and that it uses only the tabulated differences $\delta^{(n)}$; but constants of integration and of summation have to be determined. We have developed a formula for numerical differentiation which combines the simplicity of the first method with the convergence and dependence on available differences of the second. It is

$$wr_0' = \tfrac{1}{2}(r_1 - r_{-1}) - w^2\{R_0(r_1'' - r_{-1}'') + R_2(\delta_1'' - \delta_{-1}'') + R_4(\delta_1^{\mathrm{iv}} - \delta_{-1}^{\mathrm{iv}}) + \ldots .\}$$

where

$$\begin{array}{ll}
R_0 = +0{\cdot}08333\ 33333 & R_8 = +0{\cdot}00008\ 41772 \\
R_2 = -0{\cdot}00972\ 22222 & R_{10} = -0{\cdot}00001\ 89709 \\
R_4 = +0{\cdot}00183\ 53175 & R_{12} = +0{\cdot}00000\ 43526 \\
R_6 = -0{\cdot}00038\ 38736 & R_{14} = -0{\cdot}00000\ 10113 \\
 & R_{16} = +0{\cdot}00000\ 02372.
\end{array}$$

This formula has been obtained from

$$wr_0' = \tfrac{1}{2}(r_1 - r_{-1}) - \tfrac{1}{6}\Delta_0''' + \tfrac{1}{30}\Delta_0^{\mathrm{v}} - \tfrac{1}{140}\Delta_0^{\mathrm{vii}} + \ldots$$

given by Oppolzer (1880, p. 21) and Comrie (1947, table XXVII), by means of substitutions such as

$$\Delta_0''' = w^2(\delta_0' + m_2\,\delta_0''' + m_4\,\delta_0^{\mathrm{v}} + m_6\,\delta_0^{\mathrm{vii}} + \ldots)$$

and

$$\Delta_0^{\mathrm{v}} = w^2(\delta_0''' + m_2\,\delta_0^{\mathrm{v}} + m_4\,\delta_0^{\mathrm{vii}} \ldots)$$

where m_2, m_4, etc. are the coefficients in equation (24).

§ 6. Conclusion

Method 4 and Burfoot's method are superior to all the others considered. Of these two, method 4 has the advantage of being applicable also to the general ray equation. In addition, its use is somewhat less laborious and it offers the further advantages that the coefficients of the correcting terms are smaller and are available up to a very high order. These considerations apart, however, Burfoot's method is equivalent; it would, for example, undoubtedly give equal accuracy in the trace through the 'bell-field' whose results have been given in § 4.

Where primary aberrations are to be determined, Liebmann's method is still applicable, but advantage will be gained by using the formulae (9)–(16).

ACKNOWLEDGMENTS

We thank Mr. R. E. Siday and Dr. C. F. Dietrich for discussion and helpful criticism and Mr. D. K. Das Gupta for his collaboration in ray tracing through one of the lens fields.

REFERENCES

BASHFORTH, F., and ADAMS, J. C., 1883, *An Attempt to Test the Theories of Capillary Action* (Cambridge : University Press).

BOOTH, A. D., 1954, *Nature, Lond.*, **173,** 635.

BURFOOT, J. C., 1952, *Brit. J. Appl. Phys.*, **3,** 22.

CLENSHAW, C. W., and OLVER, F. W. J., 1951, *Math. Tab., Wash.*, **5,** 34.

COMRIE, L. J., 1944, *Math. Gaz.*, **28,** 90; 1947, *Interpolation and Allied Tables* (London : H.M. Stationery Office).

COSSLETT, V. E., 1946, *Introduction to Electron Optics* (Oxford : Clarendon Press).

DIETRICH, C. F., 1951, *Ph.D. Thesis* (University of London).

DUNCAN, W. J., 1948, *Phil. Mag.*, **39,** 493.

FOX, L., and GOODWIN, E. T., 1949, *Proc. Camb. Phil. Soc.*, **45,** 373.

GAUTIER, P., 1953, *J. Phys. Radium*, **14,** 524.

GLASER, W., 1941, *Z. Phys.*, **117,** 285.

GODDARD, L. S., 1944, *Proc. Phys. Soc.*, **56,** 372.

HARTREE, D. R., 1952, *Numerical Analysis* (Oxford : Clarendon Press).

HUTTER, R. G. E., 1945, *J. Appl. Phys.*, **16,** 678.

KHURGIN, YA. L., 1939, *Zh. Eksper. Teor. Fiz.*, **9,** 824.

LAUDET, M., 1953, *J. Phys. Radium*, **14,** 604.

LIEBMANN, G., 1949, *Proc. Phys. Soc.* B, **62,** 753.

LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64,** 956.

MACNAUGHTON, M. M., 1952, *Proc. Phys. Soc.* B, **65,** 590.

MALOFF, I. G., and EPSTEIN, D. W., 1938, *Electron Optics in Television* (New York McGraw-Hill).

MILNE, W. E., 1933, *Amer. Math. Mon.*, **40,** 322; 1949, *Numerical Calculus* (Princeton University Press); 1953, *Numerical Solution of Differential Equations* (New York : John Wiley).

VON OPPOLZER, T. R., 1880, *Lehrbuch zur Bahnbestimmung der Kometen und Planeten*, Band 2 (Leipzig : W. Engelmann).

PICHT, J., 1932, *Ann. Phys., Lpz.*, **15,** 926.

RICHARDSON, L. F., 1927, *Phil. Trans. Roy. Soc.*, A, **226,** 299.

STÖRMER, C., 1907, *Arch. Sci. Phys. Nat.*, **24,** 5, 113, 221.

STURROCK, P. A., 1950, *Proc. Phys. Soc.* B, **63,** 954.

TODD, J., 1950, *Math. Tab., Wash.*, **4,** 39.

A Study of Domain Structures in Alnico

By L. F. BATES and D. H. MARTIN†
University of Nottingham

MS. received 29th March 1955

Abstract. The Bitter figure technique is used to study domain structures on polycrystalline alnico specimens with coercivities ranging from 0 to 100 oersteds, in which recognizable domains are recorded. In low coercivity specimens the domains are bounded by comparatively straight walls but they show more irregular outlines in high coercivity specimens.

§ 1. Introduction

It is well known that alnico (Al 10, Ni 18, Co 12, Cu 6, Fe 54), when quenched from 1250°c in water and subsequently annealed at 600°c, develops pronounced permanent magnet properties. The coercivity increases from an oersted or so in the quenched condition to values in the neighbourhood of 300 oersteds for specimens subjected to prolonged annealing at 600°c. The effect is known to be due to the precipitation of a metallurgical phase in the form of sub-microscopic islands. The present paper records a powder deposit examination of the ferromagnetic domain structures in four specimens of alnico, denoted by S1, S2, S3, S4, in which had been developed coercivities of about 2 (quenched), 10, 40 and 100 oersteds respectively. The powder deposit technique has often been used for studies of domains in soft magnetic materials and has been described for example by Bates (1954) and Mee (1950). The patterns are formed on highly polished surfaces by fine magnetic powders deposited from colloidal suspensions, and when interpreted indicate the form of the underlying domain structure.

§ 2. Experimental Results

The four specimens of alnico were polycrystalline, with about 10 grains per mm^2. The patterns obtained clearly revealed the polycrystalline nature of the samples since the type of pattern depended markedly on the orientation of the underlying crystallite. Some crystallites formed comparatively simple patterns and their simplicity could be attributed to the presence of an easy direction of magnetization running parallel to the surface. It is this type of pattern which actually provided most of the information of interest, and the photographs recorded here are therefore mainly of this type. In many of the photographs the grain boundaries of the crystallites of major interest have been marked in ink for greater clarity in description.

† Now at Queen Mary College, University of London.

2.1. *Specimens S*1 *and S*2

Patterns obtained on these specimens of comparatively low coercivities were very similar, and examples are shown in figures 1, 2, 3 and 4.† They indicate domain structures of the type depicted in the diagram, figure 1 (*b*), that is, a system composed of domains magnetized parallel and antiparallel to one another. The domains were about 10 microns in width. It is unfortunate that we were unable to obtain surfaces free from the pitting so evident in the photographs. All the pictures were obtained when small fields were applied perpendicularly to the surfaces. These small fields, while not disturbing the domain structures to any great extent, caused deposits to form over the whole of the surfaces of alternate domains instead of at the domain walls only. Reversal of the small normal fields caused the deposits to change domains, i.e. the other set of alternate domains acquired the deposits. The pair of photographs of figure 2 illustrate this effect, which has been frequently observed in cobalt and explained by Mee (1950).

The patterns obtained on crystallites with less suitable orientations ranged from the types in figures 1–4 to the maze or honeycomb type shown in figure 5. Figure 6 shows a general view of a number of crystallites with various patterns.

Patterns like those described above bear a striking similarity to those found on crystals of cobalt (Mee 1950). The type of structure shown in figure 1 is found on surfaces of cobalt crystals which contain the easy direction of magnetization, the hexad axis. Maze or honeycomb patterns similar to the alnico pattern in figure 5 are found on surfaces of cobalt parallel to the basal plane. Both types of pattern indicate domain structures made up of ' bundles ' of needle-shaped domains, each domain being magnetized parallel or antiparallel to its length. The simple patterns, like those of figure 1, represent sections parallel to the directions of magnetization of the domains themselves. while the maze patterns of figure 5 reveal the complex subdivision of the domains at surfaces perpendicular to their directions of magnetization. The uniaxial magnetocrystalline anisotropy of cobalt provides an explanation of this type of structure in that metal, but an explanation of its appearance in alnico is not obvious. It is suggested that the strains created during the quenching process are the cause of the apparent uniaxial anisotropy in the alnico crystallites.

Magnetization processes in applied fields were observed with specimens S1 and S2. When a field was applied in a direction parallel to the domain walls, magnetization proceeded by the displacement of the walls, alternate domains expanding at the expense of the others. Such a process is recorded in figure 3. When the field was applied approximately at right angles to the lines marking the walls, no change was at first apparent. If the field was comparable with the coercivity, wall displacement began to occur, the domains magnetized slightly more favourably with respect to the magnetic field increasing in width at the expense of the others. Such a process is shown in figure 4. The absence of any re-arrangement of the structure to form a configuration including 90° walls confirms the presence of uniaxial anisotropy.

2.2. *Specimen S*3

Specimen S3 had been annealed at 600°c for six hours and had thus developed a coercivity of some 40 oersteds. Patterns obtained on it were similar to those found on specimens S1 and S2. Figures 5 and 7 show respectively grains containing a maze pattern and a simple structure. Interesting evidence that the sizes of the domains depended to some extent on magnetic history now appeared. For example, figures 7 and 8 were obtained on the same crystallite, but whereas the former was taken immediately after the heat treatment (thermal demagnetization), the latter was found after the specimen had been demagnetized by field reversals, and the domains have been decreased markedly in size.

---

† All figures are reproduced as plates at the end of this issue.

2.3. *Specimen S4*

Specimen S4 had been annealed at 600°c for 27 hours, and had thereby acquired a coercivity of about 100 oersteds. Powder patterns were obtained as before, but the domains were no longer bounded by straight walls, but had in general somewhat irregular outlines. Examples are given in figures 9–13. The reversal of the small normal applied field produced the usual movement of colloid deposits from one set of alternate domains to the other. This is illustrated in the pairs of photographs of figures 9 and 10. A study of magnetization processes in this specimen was difficult since the deposit technique lost its effectiveness in fields as large as the coercivity of this specimen. In high fields the powder particles were attracted strongly by the surface defects and the heavy deposits thus formed tended to mask the true domain structures. Nevertheless, boundary displacement seemed to occur in this specimen also.

§ 3. DISCUSSION AND CONCLUSIONS

The only previously published powder patterns obtained on precipitation-hardened alloys were those found on quenched alnico V (Alcomax) of composition 51 Fe, 24 Co, 14 Ni, 8 Al, 3 Cu (Nesbitt 1950) which in that condition exhibited a small coercivity. The patterns recorded in the present work show that recognizable domains can be observed in specimens of alnico in which coercivities of at least 100 oersteds have been developed. Both single-domain mechanisms and a wall displacement model have been considered as possible magnetization processes in alnico specimens which have maximum coercivity. For specimens with coercivities of 100 oersteds or less it is clear from the present study that the latter model is the more accurate. Even if domains with well formed walls are to be expected in the fully developed condition it is generally considered that they will be very small in order that as many as possible of the islands of free poles, which are created by the precipitation process, should be bisected by domain walls. Although a coercivity of 100 oersteds represents about a third of the maximum which can be attained using the heat treatment selected, and about a half of that value at which the most pronounced magnetic viscosity occurs (Phillips, Street and Woolley 1954), no very clear evidence of a decrease in domain size below that of the quenched condition was observed.

The most obvious change in the above patterns appeared in the shapes of the domains. In the quenched condition the domains were bounded by comparatively straight walls, while in specimens with higher coercivities there was a tendency for domains to show a more irregular outline. The increase in coercivity denotes the appearance of centres of impedance to wall movement and it is therefore not surprising that domain walls should adopt positions and orientations which depend on the local conditions.

The dependence of domain size upon magnetic history, shown in figures 7 and 8, is perhaps also to be expected. If domain walls are moved only with difficulty, a crystallite will not be able to adopt the optimum domain size spontaneously.

It would clearly be desirable to obtain pictures of domains in specimens which have been treated to produce coercivities higher than that of S4. If the domains are less readily recognizable because of their irregular shapes it becomes essential to have surfaces free from pits and other defects for convincing results to be obtained. It is this feature which has held up further progress in this direction. The severely pitted surfaces of the specimens used in the present work are perhaps due to defects in the material itself, for it was very brittle. It is hoped, however, that it will be possible to obtain better surfaces, which will permit even more definite conclusions to be drawn concerning the dependence of domain size and form upon the degree of inhomogeneity and upon magnetic history, and the work is being continued.

ACKNOWLEDGMENTS

Our thanks are due to Mr. J. H. Phillips for advice on, and arranging for, appropriate heat treatments, to Dr. A. Edwards and the Permanent Magnet Association for the alnico used in the work, and to the Electrical and Allied Industries Research Association for its continued support and permission to publish.

REFERENCES

BATES, L. F., 1954, *J. Inst. Met.*, **82,** 417.
MEE, C. D., 1950, *Proc. Phys. Soc.* A, **63,** 922.
NESBITT, E. A., 1950, *Amer. J. Phys.*, **21,** 879.
PHILLIPS, J. H., STREET, R., and WOOLLEY, J. C., 1954, *Phil. Mag.* (7), **45,** 505

The Sounds of Impact of a Solid on a Liquid Surface

By E. G. RICHARDSON
King's College, Newcastle upon Tyne

MS. received 22*nd November* 1954 *and in amended form* 21*st March* 1955

Abstract. The sounds, recorded on a microphone in air and on a hydrophone beneath the liquid, when a solid body strikes the surface are recorded and correlated with cine-photographs of the cavity formed on entry. Measurements of base pressure on a projectile forming a cavity are also made. It appears that the major contribution to the sound comes from pulsations of the cavity.

---

§ 1. Introduction

In an earlier paper hereafter referred to as I the author (1948) described experiments on the cavities formed when solid bodies strike a liquid surface. Since then a number of other papers have appeared on the subject, in particular those of Gilbarg and Anderson (1948), May and Woodhull (1948), May (1951, 1952). All these papers have been concerned with the dynamics of water entry. It was thought by the author that an investigation of the sounds on impact might be of interest.

This was indeed the problem which, judging from the title of his paper, Mallock (1918), an early investigator, set himself but did not in fact solve. On the other hand Minnaert (1933) and others have studied the sounds of splashes and bubbles formed in natural water courses, while Meyer and Tamm (1939) have studied the natural frequencies and damping factors of small bubbles formed *under* water.

The scheme of the experiments to be described was to drop or shoot solid spheres from moderate heights into a tank of liquid, photograph the cavity and splash so formed on a cine-film while records of the sounds picked up by a microphone above and a hydrophone below the liquid were made, then to attempt to correlate the two.

§ 2. Evolution of Entry Cavities

When an object hits the liquid vertically it sends up a splash and cleaves the water as it enters to form a cavity which, unless the missile has supersonic velocity, quickly fills with air. After its initial horizontal impulse, the cavity wall decelerates, reverses direction and seals at the surface due to the falling down of the liquid originally thrown into the splash and, later, at a depth due to the collapse of the cavity walls.

The whole phenomenon may be shown (I) to scale on a Froude basis, i.e. if V is the entry speed, D the diameter of the obstacle—in most experiments a sphere—the evolution of the cavities and their scaling is governed by the constancy of V^2/gD. From the films one can derive the maximum volume of the cavity which occurs some time after surface seal at about the instant when, in fact, the 'lid' formed at the surface tends to be pulled down into the liquid. Thereafter deep closure takes place, leaving from half to one quarter of the cavity attached to the projectile. The remainder pulsates, eventually splitting up further into pieces which rise and break surface.

§ 3. SOURCES OF SOUND

Among possible sources of the noise picked up by microphone or hydrophone, we may consider:

(1) the 'slap' on the water surface, due to the 'rigidity' of the latter;

(2) natural vibrations of the projectile;

(3) the splash of water, thrown up from the surface;

(4) pulsations of the cavity volume, due to changes of pressure inside it, accompanied by motion of the walls of the cavity;

(5) oscillations of the air inside the cavity in resonator fashion, particularly when the roof is open to the atmosphere, or when bubbles break surface.

Of these, (3) and (5) would be most apparent in the air, (4) in the water, while (1) and (2) might affect both microphone and hydrophone.

While (1) and (3) may be rather indeterminate in frequency, (2) can be calculated (for a sphere) and sought in the noise spectrum; (1) and (2) start at instant of touch-down, (3) rather late. The remaining two sources both involve pressure fluctuations inside the cavity, (4) in the gas as a whole, (5) with differences of amplitude, possibly also of phase, over the gaseous volume—Helmholtz resonator or organ-pipe oscillations—without involving any motion of the walls.

In some of the photographs there is evidence of another type of vibration in the cavity, i.e. those of high order bell-like oscillations of the cavity wall, for which —in a sphere—Rayleigh (1879) gives the formula

$$\omega^2=(n+1)(n-1)(n+2)\{\sigma/\rho a^3\}$$

where ω is the pulsatance, a the radius, σ surface tension and ρ density of the liquid, n being an integer greater than unity.

The evidence for these is the regular 'fluting' of the cavity walls, which probably indicate nodal lines of longitude, though such a high-order oscillation would not radiate to a distance and furthermore involves no change of size of the cavity, so that it need scarcely be considered seriously in this connection. If such a 'fluted' cavity is photographed from above the appearance, in a time exposure, is somewhat as figure 1 shows, with dark segments between the nodes indicating the existence of oscillation.

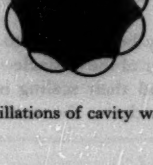


Figure 1. Flexural oscillations of cavity walls, as seen from above.

Before taking measurements of the sound radiated from the cavity into the surrounding air and liquid, it was decided to attempt to measure the pressure fluctuations within the presumed source of sound itself, by fitting the plane base of an ogive-nosed projectile with a thin metal diaphragm and fitting inside the shell a self-contained recorder of the vibrations of the diaphragm. The latter was a D.V.L. scratch recorder shown in section in figure 2. The motion of the diaphragm Z, magnified by a system of levers V is scratched by a diamond U on a stainless steel ring T rotated through the friction wheel R by clockwork and set in action at the instant of dropping the body on the water. The records were subsequently projected in an episcope to give a magnified trace on paper. Figure 3 shows three typical records and details of eight successful drops are given in the table.

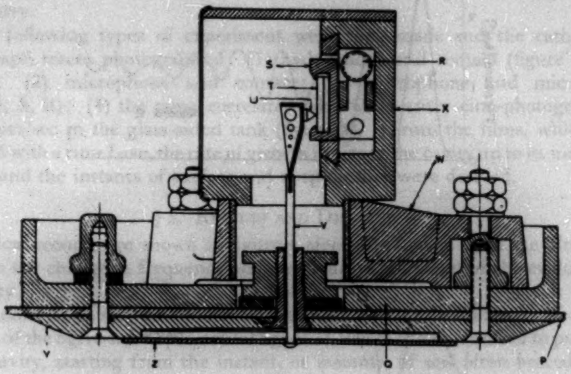


Figure 2. Scratch recorder; the vibrations of the membrane Z are inscribed by the diamond U on the moving stainless steel strip T.

No.	h (ft)	V (ft)	F	a_1 to a_{10}	δ
1	4	16	60 to 90	11 to 0·2	0·35 to 0·6
2	4	20	50 to 80	11 to 0·4	0·27 to 0·5
3	4	19	60	11 to 0·2	0·40
4	3	14	60	7 to 0·2	0·20
5	3	14	70	2 to 0·2	0·27
6	4	18	40	—	—
7	3	14	50	4·5 to 0·4	0·25
8	3	14	60	4·5 to 0·4	0·25

The second column gives the height of drop, the third the entry velocity, the fourth frequencies of pressure oscillations (deduced from the 'wavelengths' on the record), the fifth their amplitudes reckoned from the initial and the tenth oscillation, the sixth the decrement.

Tenth-second time marks were made on the pressure records by a clockwork device, and a standard fork gave a record on one beam of the oscillograph used later, but when two beams were in use the fork record was made immediately before and after the sound record.

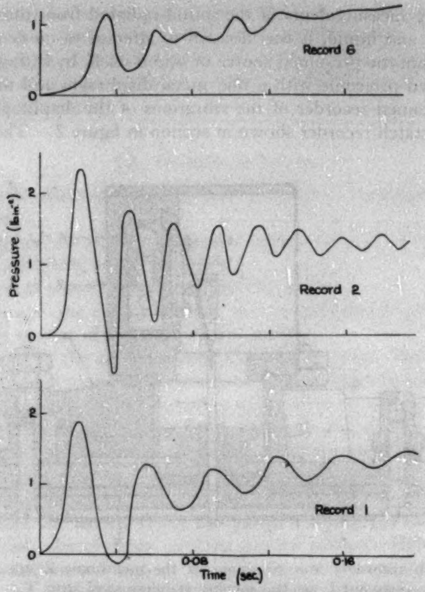


Figure 3. Base-pressure records in cavities formed by ogive-headed projectile.

§ 4. EQUIPMENT FOR SOUND MEASUREMENTS

For the airborne sound a microphone and amplifier was used so that the wave form could be exhibited on a split-beam cathode-ray oscillograph. For the water-borne sound small electromagnetic and crystal hydrophones with high natural frequencies were constructed and sealed with picsin to prevent entry of water. The record of either of these was also amplified and shown on the cathode-ray oscillograph. No obvious differences appeared in their records, when they were simultaneously employed.

Spheres of steel (ball bearings) or aluminium of diameter 1 in., 1½ in., 3 in. and 5 in. and a 3 in. sphere of varnished wood were used and dropped or catapulted from heights up to 8 ft into either a glass-walled tank or a larger tank.

The latter, of dimensions 4 ft × 4 ft × 12 ft, was more suitable for this work since its size allowed both microphone and hydrophone to be set at equivalent distances (1 ft and 4 ft respectively) from the point of impact. (The microphone was protected from splash by a muslin hood.) Moreover, the tank was lined with arrays of rubber prisms ('fafnir') to reduce reflections from the walls. It is known that reflections affect the size and so the natural frequencies of cavities formed in small rigid-walled tanks.

For calibration purposes an underwater loudspeaker was used as a source and the hydrophones set successively at the same point in the tank as another probe which had itself been calibrated in a tank belonging to the Admiralty.

The hydrophones were flat within 5 dB from 50 to 1000 c/s. The electromagnetic type had a slight eminence near 400 c/s and the crystal type several peaks above 2 kc/s.

To disentangle the possible sources of sound (§3) it was desirable to relate the onset of the sound to the instant of impact. To this end a pair of foil contacts could be arranged just at the water surface so that the 'touch-down' by the sphere applied a steady potential to one beam without appreciably interfering with the water entry.

The following types of experiment were then made and the cathode-ray oscillograph traces photographed: (1) hydrophone and contact (figure 4, C, D (Plate)); (2) microphone and contact; (3) hydrophone and microphone (figure 4, A, B); (4) the same correlated with water-entry cine-photography at 200 frames/sec in the glass-sided tank (figure 4). From the films, which were provided with a time base, the rate of growth in size of the cavity up to its maximum volume and the instants of surface and deep closure were derived.

§ 5. RESULTS AND DISCUSSION

Typical records are shown in figure 4 alongside frames from the film which illustrate the change in frequency and amplitude in the light of the evolutions of the cavity.

The chief frequency detectable in the oscillograms is an oscillation of frequency like that of the ogive-nosed projectile (figure 3), which can be ascribed to pulsations of the cavity, starting from the instant, or instants, of seal often preceded by a more intense short-lived oscillation which represents the 'slap' on closure of the cavity.

Minnaert (1933) and Smith (1935) obtained a formula for the frequency of pulsation of a sphere, by equating kinetic to potential energy in the vibrating gas, which was verified, for small spherical bubbles, by the former and by Meyer and Tamm (1939). For air bubbles in water the latter find the frequency f to be inversely as the radius a with the formula $fa=326$. These cavities are not spherical but inasmuch as the theory only involves change in volume, we may still expect this formula to apply approximately if we take the mean radius of the volume, deduced from the cavity volume at the instant of closure.

If v_m stands for the maximum volume (reached shortly after surface seal), dimensional analysis indicates $v_{\mathrm{m}}D^{-3}\propto VD^{-1/2}g^{-1/2}$ or $v_{\mathrm{m}}\propto D^{5/2}$. We should then have $f_0^{-1}\propto v_{\mathrm{m}}^{1/3}\propto V^{1/3}D^{5/6}\simeq V^{1/3}D$ (f_0 stands for the lowest frequency occurring at the beginning of the closure (figures 3 and 4).

When deep closure occurs, the frequency of the main cavity will rise and this rise will continue as the two portions break into further fragments. On some records indeed beats can be seen in places, probably indicating a difference in frequency between the broken portions.

The formula last quoted is justified by the results, as figure 5 shows.

The further deep sealing and fracture of the cavity into pieces causes repercussions of the sound at higher frequencies along the record (figure 4 A) or a gradual rise in the base pressure (figure 3). These evolutions do not in fact, produce catastrophic rises in natural frequency for, since the latter is proportional to $v^{1/3}$, a splitting of the whole cavity into two equal portions—which is more or less what actually takes places—only raises the pitch in the proportion 4:5. The greatest change which has been observed in one pressure trace is 2:3.

A rather rapidly damped high-frequency underwater sound can be detected (figures 4 C and D), if sufficient gain is used, at the instant of impact. This, more prominent in the records with the metal than the wooden balls, is ascribed to natural vibrations of the solid excited on impact, such as one gets by tapping with a hammer.

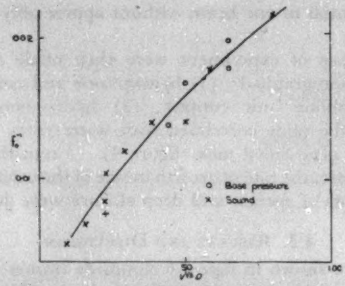


Figure 5. Reciprocal of initial cavity frequency plotted against the product of projectile diameter and the cube root of impact velocity.

As to the intensity of the sound picked up, the force of impact is, as is shown in I proportional to DV^2. If the projectile can then be considered to change its velocity suddenly to a smaller value V_1, due to the 'added mass' of liquid, it covers this distance V_1 in the succeeding second. We may consider then the sound radiated into the water in unit time proportional to V^2V_1 or to kV^3 where k is a factor depending on the added mass and, as measured for various types of 'head', given in the fifth column of table 2 of I. Thus an ogive head should radiate five sixths of the sound of a hemisphere, two thirds of that of a flat disc, other conditions being equal.

It must be stated, however, that in spite of this high exponent in the relation between sound intensity and impact velocity, the intensity of sound was not so 'repeatable' as the frequency, even with a projectile of the same diameter and density. Sometimes 'soft impacts' were recorded for which no explanation is offered. May (1951) points out that impacts at low speeds are not always reproducible.

In figure 6 is shown the decrement δ of the principal oscillation measured (*a*) for the ogive projectile from the pressure record, (*b*) for the spheres from the sound records. The greater damping in the former may be instrumental, but both seem to be proportional to frequency directly. Meyer and Tamm (1939) and Exner and Hampe (1953) point out that on the assumption of spherical radiation and an inviscid liquid, radiation resistance would require δ to be 0·045 for all frequencies, but in fact they found the damping of their small bubbles to rise nearly in proportion to the natural frequency. Some of this effect may be due to the heat conductivity but it is small at low frequencies, being a function of $2a(\pi f\kappa)^{1/2}$ (Pfriem 1940). Viscous damping should be in proportion to frequency. A similar feature is apparent in these results but the values of the decrement are very much larger than is deduced from the results on the spherical bubbles extrapolated to low frequencies. This is not unexpected, considering the transfer of energy into turbulent motion by the vibration, change of size and break-up of the cavities.

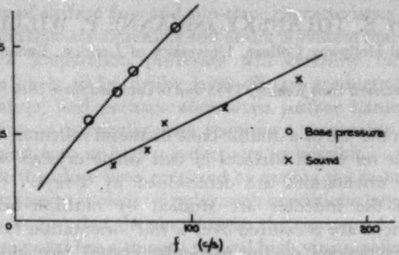


Figure 6. Decrement of cavity pulsations plotted against frequency.

It is interesting to note, from a study of the pressure records of figure 3, that as the frequency of the cavity vibrations rises in time (see above) so does the decrement. Thus on the first record, as f goes from 60 to 90 c/s in the length shown, δ rises from 0·45 to 0·6 sec^{-1}.

Some sound records and cavity photographs were also obtained when the 1 in. and $1\frac{1}{2}$ in. balls were shot into glycerine. For economy's sake these were done in a small tank (1 ft × 1 ft × 4 ft) and were not free from tank oscillations at the instant of impact. These however had died away by the time the cavity sealed. The frequencies of the ensuing pulsations were the same as those recorded for equal conditions in water but their decrements were two to three times those in water, a feature noted by Meyer and Tamm (1939) for small bubbles in glycerine.

ACKNOWLEDGMENTS

The author wishes to thank the Royal Naval Scientific Service for assistance in providing apparatus and making possible the services of a Technical Assistant, Mr. R. F. Fraser, to whom he is also indebted.

REFERENCES

EXNER, M. L., and HAMPE, W., 1953, *Acustica*, **3**, 49.
GILBARG, D., and ANDERSON, R. A., 1948, *J. Appl. Phys.*, **19**, 127.
MALLOCK, A., 1918, *Proc. Roy. Soc.* A, **95**, 138.
MAY, A., 1951, *J. Appl. Phys.*, **22**, 1912 ; 1952, *Ibid.*, **23**, 1362.
MAY, A., and WOODHULL, J., C., 1948, *J. Appl. Phys.*, **19**, 1109.
MEYER, E., and TAMM, K., 1939, *Akust. Z.*, **4**, 145.
MINNAERT, M., 1933, *Phil. Mag.*, **16**, 235.
PFRIEM, H., 1940, *Akust. Z.*, **5**, 202.
RAYLEIGH, LORD, 1879, *Proc. Roy. Soc.*, **29**, 71.
RICHARDSON, E. G., 1948, *Proc. Phys. Soc.*, **61**, 352.
SMITH, F. D., 1935, *Phil. Mag.*, **19**, 1147.

Directional Hardness Studies on Single Crystals of Tin and Bismuth

BY S. TOLANSKY AND ANNE P. WILLIAMS
Royal Holloway College, University of London, Egham, Surrey

MS. received 18*th January* 1955 *and in amended form* 16*th March* 1955

Abstract. With the use of a double-cone diamond indenter, directional hardness studies are made on virgin surfaces of cast single crystals of Sn and Bi. The crystallographic orientations are determined by x-rays. . Surface flow effects associated with the indenter are studied by multiple-beam interferometry. Hardness variations are measured over a 180° orientation range on each surface studied and they depend on the particular crystal face indented. Thus on a tin (311) face the ratio of maximum to minimum hardness goes through the range 1·6 : 1 whilst on a (554) face of bismuth the range is as great as 4·5 : 1.

It is demonstrated by interferometry that slip mechanism plays a primary part in producing directional variations in hardness.

---

§ 1. INTRODUCTION

ALTHOUGH the existence of variations in directional hardness for several metal crystals has long been known, yet little work in this field has been published. This is probably associated with the fact that the older and still widely used hardness measuring devices are not well suited for measuring directional effects. It can be anticipated that hardness in a metal crystal will be a function of crystallographic direction. O'Neill (1923) reported that ball indentations on single crystals of aluminium produced elliptic impressions due to hardness variations in different directions. Schultz and Hanemann (1941) also found slight asymmetries in the indents on aluminium crystals, when a Vickers diamond pyramid was used, and directional variations of hardness up to 7% were found by rotating the crystal relative to the indenting tool. Meincke (1950) described scratch tests on single crystals of Al, Zn and Cu and later (1951) reported measurements made with a diamond chisel. The scratch test data reported by Meincke gave surprisingly large variations, for according to him both in Zn and Cu the hardness on one face was no less than 2½ times that on the other.

In no case quoted above was a study made of the flow pattern of the material surrounding the indent. However, Tolansky and Nickols (1949, 1952) by applying multiple-beam interferometric methods to the examination of surface flow around Vickers pyramid indents on single crystals of tin revealed the existence of curious flow anomalies closely related to crystallographic directions. The indents instead of being square-shaped had two opposite sides convex with the other two concave. There was also marked surface pile-up in the direction of the *c*-axis of the crystal, irrespective of the orientation of the sides of the indent, and to a lesser, but no less striking degree, sink-in in a direction at right angles to the pile-up.

In view of this crystallographic flow effect the existence of a directional hardness variation was clear. Study has therefore been carried out on a number of tin and bismuth crystals. This has aimed at measuring directional variations of indentation hardness on known (arbitrary) crystal faces and at the same time an interferometric study of surface flow and slip-trace characteristics has been made.

Tin and bismuth crystals are extremely soft and it is evident that any mechanical polishing or surface preparation processes will certainly introduce serious falsities. There are likely to be Beilby layer effects, production of micro-slip, possible twin formation, and perhaps anomalous surface hardening. For the interferometric studies in particular, one requires a reasonably flat surface, as smooth as possible and with as high a reflectivity as possible.

The crystals have therefore been prepared by casting the metal on to polished glass, the melt being slowly cooled. Under such conditions large single crystals can be grown. It is found that the tin takes on the contour of the glass surface very closely and on removal has a smooth natural high virgin polish, ideally suited to interferometric study. A micro-hardness indentation on such a surface will not be falsified by mechanical surface preparation. In bismuth contouring is less perfect, though adequate.

The drawback to this technique lies in the absence of control over the crystal direction lying in the surface. It would have been preferable to have made observations on natural crystal facets with selected low indices. This we have been unable to achieve yet and have had to be content with arbitrary planes whose orientations we have, however, determined by x-ray methods. The surfaces examined are, however, virgin surfaces and any observed hardness variations are undoubtedly truly crystallographic. Despite repeated attempts which include seeding, pouring from the melt, and electrolytic deposition, we have failed entirely to secure satisfactory single crystals of tin exhibiting plane crystallographic faces of a size sufficiently large to accept an indentation tool and have found it necessary to make studies on the arbitrary crystal faces arising from the casting method. We have sought the expert metallurgical advice available at the Tin Research Institute, but as yet it has not been possible to secure a suitable crystal with sufficiently good natural facets. The single crystals produced have very curved faces and these are clearly not suitable.

The beautiful bismuth crystals which can be grown easily from the melt never seem to exhibit flat faces.

§2. EVALUATION OF DIRECTIONAL HARDNESS

There exists a wide variety of hardness tests (O'Neill 1934) and it is agreed that they do not all measure the same physical quantity. The more widely used methods involve either an indentation or a scratch. In common use for indentation hardness studies are (*a*) the ball, (*b*) the right-circular Rockwell diamond cone (*c*) the square-based Vickers diamond pyramid, and (*d*) the Knoop elongated diamond rhomb. More recently (*e*) the diamond chisel and (*f*) the diamond double-cone have been advocated.

We have considered the characteristics of these various methods, primarily as to the effects produced on a crystal or other surface which has variations in hardness in different directions. Scratching by a diamond (see Meincke 1950) does indeed reveal directional hardness, but the technique is notoriously difficult to apply and liable to serious systematic error. According to Bierbaum the relatively strong corner of a diamond cube is the best scratching point. The scratches formed are always narrow, and the width, which is the required dimension, is difficult to measure accurately, partly because of the scale and partly of irregularity and distortion of the edges due to plough-up of material. Much more serious is the impossibility of recognizing and allowing for recovery effects. We have established that long narrow *indents* can reveal exceptionally big recovery effects in their *width* dimensions and, as scratches have somewhat related characteristics, it may well be that errors of even hundreds per cent can arise through recovery defects. Although undoubtedly capable of revealing qualitative differences we reject the scratch method as unsatisfactory for quantitative measurement.

The ball indentation, and also the cone, produces slightly elliptical impressions on an anisotropic material but these are not easily amenable either to exact measurement or to clear interpretation. Non-symmetrical indents which appear on anisotropic materials when using a Vickers diamond square-based pyramid are very difficult to interpret quantitatively, although they certainly do reveal qualitative directional differences (Tolansky and Nickols 1949, 1952). The diamond chisel advocated by Meincke (1951) appears to us to be even more unsatisfactory than the scratch test. With a chisel the width of a long narrow indent is measured and in addition to recovery defects there are serious end-effects. Furthermore, the method is restricted to indents of relatively large dimension, for clearly the smaller the chisel the more serious are the end-effects. Consequently heavy loads are often required, involving a strong massive machine.

There remain, for consideration, the Knoop rhomb (Peters and Knoop 1940) and the double-cone (Grodzinski 1952). The Knoop rhomb, favoured in U.S.A., is an elongated pyramid which produces a long narrow rhomb-shaped impression which gives a much closer approach to a directional hardness figure than any of the above methods. It still retains some of the characteristics of the square-based pyramid. It is reasonable to expect that recovery will affect the minor diagonal rather than the major diagonal and this is an important advantage.

The double-cone, designed by P. Grodzinski, appears to us to be the best instrument for our purpose. Although recently described it is an improved development of a much earlier proposal. Indentation measurements with a double-conical knife were first described long ago by Middleberg (1886), and Frémont (1924) used a 5 cm double conical wheel of V-shaped profile.

The double-cone, although cut from a single diamond, may be regarded as the equivalent of two right circular cones placed with bases in contact and vertices opposing, such that the curved surfaces meet to include an angle exceeding 120°. Such an indenting tool has a strong edge and when a load is applied as shown schematically in figure 1 the result is the production of an elongated, shallow, boat-shaped impression. The theory of the instrument has been given by Grodzinski (1951) who has calculated the area of the indent, assuming no recovery, and thus derived an expression for double-cone indentation hardness analogous to that of the widely used Vickers pyramid hardness. The area of the impression is, closely enough, $A = \frac{1}{8}l^3 \tan \frac{1}{2}\alpha / R$ in which R is the radius of the cone base, $\frac{1}{2}\alpha$ the angle between the normal to the plane being indented and the cone, as shown in figure 1, and l the length of the indent. The depth is approximately $l^2/8R$ and the width $\frac{1}{4}l^2 \tan \frac{1}{2}\alpha/R$. For our indenter $\alpha = 154°$.

For any particular indenter R and α can be regarded as constants and if a load of P kg has been employed to produce the indent then, since indentation hardness is defined as P/A, the hardness is given by CP/l^3 in which C is the instrumental constant $6R/\tan \frac{1}{2}\alpha$. In accordance with standard practice the hardness number is given in kg mm^{-2}.

Grodzinski has established, by measurement on a variety of materials with a very wide range of hardness, that the double-cone hardness, as defined above, correlates closely with the corresponding Vickers pyramid hardness. Although this is satisfactory, in that it permits correlation of double-cone hardness figures with the widely accepted Vickers pyramid criteria, yet in fact, even if there were no systematic correlation, the double-cone would still be valid for revealing directional effects on any given material.

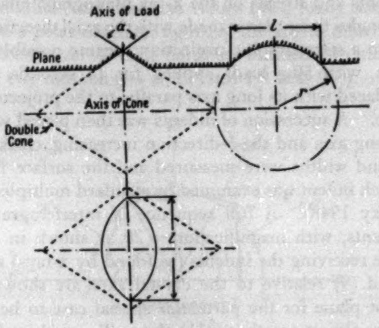


Figure 1.

We have measured the double-cone hardness of pure polycrystalline tin, (purity better than 99·995%, supplied by the Tin Research Institute) at room temperature with a 50 g load and find the value 3·9 (tin is of course very soft). It is eminently satisfactory that this agrees exactly with the best available Vickers pyramid value (3·9 at 20°c) reported by Homer and Plummer (1938), who made comprehensive measurements over the range 0° to 100°c.

This agreement gives confidence in the numerical values we quote later. Attention is drawn to the fact that mechanically worked tin has been reported (Perryman 1950) to have a substantially higher hardness (7·7), as is only to be expected, and this indicates how dangerous mechanical polishing of surfaces would be in these experiments.

We have been able, for the first time for micro-indents, to measure the depths as well as the widths and lengths of the indents. Furthermore, it has been established that recovery effects can be very marked both on the depth and width, but evidence from the behaviour of surface distortions when examining both soft and hard materials shows that the length dimension hardly appears to be affected by recovery. It is therefore considered that any observed variations in length of the indent with direction (the load being kept constant) do give a true measure of variations in directional hardness. Since the length of the indent in a given direction can be altered by varying the load the instrument is suited to micro-hardness testing and the indent size can be made sufficiently small to permit a large number of indents to be well spaced even on a small crystal face, and so placed as to escape distortion from neighbours. This is an important practical advantage possessed by the double-cone.

§ 3. MEASUREMENTS : TIN

A double-cone indenter has been adapted to fit the standard micro-hardness indenting attachment of the Vickers projection microscope. Indents have been made on selected tin crystals. In each case the directions of the crystal axes relative to the plane face receiving the indents was determined by x-ray methods. The x-ray measurements were made by us at Birkbeck College through the courtesy of Professor J. D. Bernal. The crystal was oscillated through successive 15° ranges, the spots and streaks on the x-ray photograph enabling identification of planes. The angles these planes made with the axial directions were calculated so that location on a stereographic projection became possible.

Micro-indents, with 50 g loads applied for 15 seconds, were made. The first indent was placed with its long axis parallel to the projection of the *b*-axis on the crystal surface. A succession of indents was then placed with angles between the indentation long axis and the *b*-direction increasing, clockwise, by 10° up to 180°. Lengths and widths were measured and the surface flow and distortion associated with each indent was examined by standard multiple-beam interference methods (Tolansky 1948). A full sequence of interferograms of the regions around these indents, with magnification × 20, is shown in Plates II and III.

The cast plane receiving the indents (as found by x-rays) and the orientation of indents (1) and (9) relative to the crystal axes are shown schematically in Plate I. The cast plane for the particular special case to be discussed here in detail is found to be almost exactly a (311) plane. To avoid confusing the diagram, only the position of indents (1) and (9) are shown. These are respectively parallel to the projections of the *b*- and *c*-axes in the (311) plane. The remaining indents are oriented in succession following each other at 10° intervals.

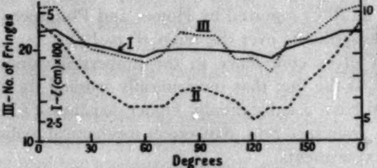


Figure 2.

For each indent it is found that the three dimensions, length, width and depth, vary with the direction and these are plotted together in figure 2; curve I shows length, curve II width and curve III depth, obtained by counting fringes within the indents.

There is fairly close parallelism between the three dimensional variations. The maximum width dimension (minor axis) cannot be considered to have much significance since we find that there are width variations along the length of the indent which are closely associated with the directions of pile-up and sink-in, as revealed by the fringes. These variations are connected with recovery and complicate the pattern.

The maximum depth measurements are of much interest in that they reveal a considerable recovery in depth. The fringe orders closely enough follow a $\sqrt{n}$ rule, characteristic of Newton's rings, thus a reliable radius of curvature along the long axis can be computed. The radius approximates to 4 mm. However, both interference and microscope measurements made on the diamond double-cone itself show that the diamond radius of curvature is only half this, i.e. 2 mm. Thus there has been practically a 50% recovery in depth. This is surprising when the shallowness of the indent is realized. (The diamond itself has certainly not flattened by being impressed on soft tin.)

The directional double-cone hardness values (50 g load) calculated from the lengths are shown on a polar diagram, figure 3, the angles referring to the *b*-axis direction which is taken as zero. The hardness exhibits two maxima (2·1 at 60° and 2·2 at 130°) and two minima (1·4 at 10° and 1·9 near 90°).

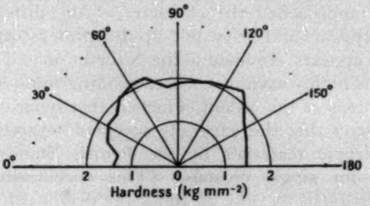


Figure 3.

The two maxima and two minima are separated by the same amount. The extreme values differ by about 60%, the ratio being 1·6 : 1. This of course applies specifically to this (311) plane only. A notable feature is that for this single crystal even the maximum value 2·2 is appreciably less than the value 3·9 which we and others find for polycrystalline tin. This will be accounted for later by considerations of the restriction imposed on slip in the polycrystalline aggregate.

Several crystals have been studied and all exhibit similar polar diagrams but the particular values for the maxima and minima are determined by the accidental orientation of the plane receiving the indents. Nevertheless, an approximately quantitative explanation can be made for any particular crystal in terms of slip directions.

§4. THE SURFACE FLOW

Surface flow directions have been identified both by Fizeau fringes and by fringes of equal chromatic order (Tolansky 1948). Consider the first indent, Plate II : surface flow is symmetrical, with hillocks about one light wave high piled up on either side. Within the hillock regions fringe discontinuities reveal the existence of slip-band formation. The slip traces can also be seen by direct micro-photography, but only the fringe pattern gives a measure of the heights. The traces observed are parallel to the long axis of the indent, are some 3 microns apart and the heights approximate to only 200 Å.

On following through the sequence of indents the direction of piling-up in all cases remains the same, although the amount decreases progressively from indent 1 to 6.

In each case the slip traces maintain the same direction as for indent 1, but the traces become more widely spaced and tend to increase in height, although there is some uncertainty in determining this because of the rapid variation in fringe dispersion (high dispersion is needed to show the small step heights).

Using a N.S.E.W. description, the following holds. The E and W pile-ups for indents 1 to 5 have associated with them N and S slight sink in, exactly in accord with the observations made with a pyramid by Tolansky and Nickols. On indents 2 and 9 (those which give minima for hardness), the pile-up maxima are in accord. The hardness maxima occur between these positions. Thus it follows that maximum hardness arises when there is greatest interference by the indent with the slip systems. As this will be more marked in a polycrystalline aggregate, this confirms that one can anticipate a higher hardness for polycrystals than for single crystals, as is indeed the case.

As indent 9 is approached the character of the distortion changes. The original pile-up disappears and a new pile-up direction associated with the second hardness minimum appears, confined to the N direction. This pile-up is 1½ light waves high, and it is highly asymmetrical, extending to no less than 14 times the length of the short axis of the indent, which is that lying in the N.S. direction. This again emphasizes that the generally accepted separation between indents, prescribed for isotropic materials by the British Standards specification, is totally inadequate for single crystals. (This specification accepts as valid pyramid indents separated by only 2½ times the diagonal.)

On following through successive interferograms 9–18 the original pile-up direction once more becomes re-established.

We have analysed in detail here only one of the many series of observations made, each on a different crystallographic face. In all cases there is an analogous behaviour, but with numerical data for the polar diagram being specific to the particular face.

In every case the fringe pattern behaviour relative to directions of axes and slip trace directions is consistently uniform, hence there is little point in multiplying illustrations by including observations made on other faces.

§ 5. BISMUTH

Although we have hopes of developing a method of growing good bismuth crystals, we are as yet obliged to resort to the casting technique to obtain virgin surfaces adequately smooth for interferometry. We find that in bismuth ('Chempur' quality) the effect of directional slip is even more striking than in tin since the number of slip systems which appear to operate is more limited than for tin crystals. On one crystal a change of hardness from 1·5 to 6·8 was observed on one face. Plate IV shows the interferograms of the two corresponding indents made in these two directions, both with 50 g loads. On this crystal face only one slip plane was suitably orientated to permit slip to take place. The first indent shows the position of minimum hardness, i.e. there is maximum flow on this slip plane. The slip traces are clearly revealed by the fringes which show discontinuities which mean that the maximum slip step is 200 Å, and as the distance from the impression is increased the steps become only 50 Å high. The second indent shown was made with the cone axis in a direction at 90° to the first, with the same applied load of 50 g. This lies in the direction least favourable for slip and is therefore the position of maximum hardness. This is not due to any interference of slip systems acting, but just because of the difficulty of slip on any system taking place. The fringes show that no piling up has taken place in any direction.

As in the case of the tin crystals the crystallographic orientation of this face has been determined. It is near enough (554), i.e. not very far removed from the important (111) face.

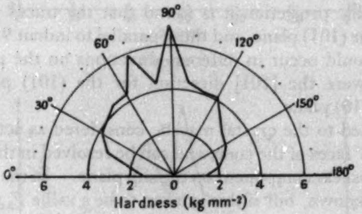


Figure 4.

The directional hardness variations found for this face, by making systematic indents over 180° orientation, are shown on a polar diagram in figure 4. The ratio of the extreme values is no less than 4·5 : 1. This surprisingly large ratio is almost twice as big even as the extreme ratio reported by Meincke for Cu and Zn with the scratch test (which we regard as suspect). Small twins were sometimes observed on bismuth crystals at the edges of the impression when the indenter was placed in certain directions. The occurrence of these, however, did in no case affect the smoothness of the hardness curve obtained.

§ 6. CONCLUSION

It seems clear that the slip mechanism plays a primary role in leading to directional hardness variations. It is suggested that our observed directional anisotropy is mainly due to the change in orientation of the indenter relative to the active slip plane and slip direction. The forces acting on the crystal during indentation can be considered to operate normally to the face of the indenter. The resolved shear stress along the slip plane due to this force has been calculated and will form the subject of a separate report. The length of indentation can thus be regarded as a measure of ease of slip, a larger indentation appearing when the resolved shear stress is high, a smaller when it is low.

For this reason one can expect quite different hardness ratios for single crystals and polycrystal masses of the same material.

APPENDIX

Crystallographic Considerations

By Anne P. Williams

The traces beside the indentations are traces of the planes on which slip has occurred in the interior of the crystal to relieve the stress applied by the indenter. The indices of the slip planes have been determined from x-ray data, the possible slip planes being known from values given by Schmid and Boas (1935). For tin the possible planes and directions are the systems (100), [001]; (101), [101]; (121), [101]; (100), [001]. The system (100), [101] is given with a query by Barrett (1943). For bismuth only the (111) planes and [101] directions are possible.

For the tin crystal the slip traces observed are parallel to indents 1 and 9. From a stereographic projection it is found that the traces parallel to indent 1 are due to slip on the (101) plane, and those parallel to indent 9 to slip on the (010) plane. The slip could occur in different directions on the planes but the most suitably oriented were the $[10\bar{1}]$ direction for the (101) plane and the [101] direction for the (010) plane.

The force applied to the crystal may be considered as acting in directions at right angles to the V faces of the cone, and can be resolved in the slip plane and slip direction giving the shear component on the slip plane. In the test the actual value of the force is not known, but if this is taken to be a value F, then $F \cos\phi \cos\lambda$ is this shear component, where ϕ is the angle between F and the normal to the plane and λ the angle between F and the slip direction (Barrett 1943).

Daniels and Dunn (1949) have made such an analysis of directional tests on crystals of silicon ferrite and zinc. The correction factor for the presence of a constraining force on the slip due to the indenter which they calculated has been applied to this work, and the resultant stress is $F \cos\phi \cos\lambda \cos\psi$ where ψ is the angle between the direction of indentation and an axis in the slip plane at right angles to the slip direction.

Now the value of F which depends on the applied load may be taken as a constant factor and the value of $\cos\phi \cos\lambda \cos\psi$ will change as the indenter is placed in the different directions on the surface. In this case the slip plane is known and the values of ϕ, λ, ψ may be obtained by the method of Daniels and Dunn. The positions of the force directions and slip planes and directions are plotted on successive stereograms and the values of $\cos\phi \cos\lambda \cos\psi$ for each position of the indenter calculated from measurements of ϕ, λ and ψ.

Figure 5 shows how the calculated curve compares with the measured length and hardness curves. When the stress applied to the slip plane is greatest this agrees with the maximum length and the minimum position of hardness. The variation agrees in range, i.e. the values for the (101) plane operate from 0–60° and 120–180° and the (010) in the intervening region, and in turn with the observed changes in pattern of fringes. The curve for the slip plane $(\bar{1}21)$ has been included in figure 5, as Barrett queries (100), [101], to show that agreement in range is also given by this plane. It was not possible to see the slip traces clearly in this direction to distinguish between the two possibilities.

Thus the variation in directional hardness may be explained qualitatively by the change in shear stress on the slip planes. For other tin crystals the slip systems (110), [001] and (121), [101] have been observed to operate. Similar results are found for bismuth (figure 6) showing the curves obtained for this crystal. The first indent, with maximum deformation, is at 0°.

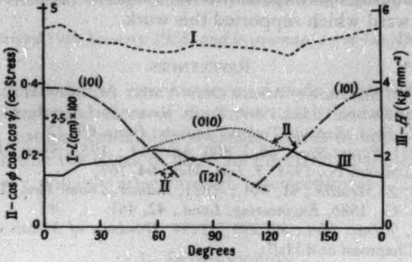


Figure 5.

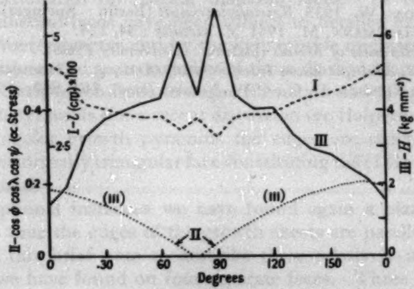


Figure 6.

For tin the yield stress is 1·4 ton in^{-2} or 2·2 kg mm^{-2} (Tabor 1951). This is approximately equal to the mean of the hardness numbers observed by Tolansky and Williams on several crystals. The yield point of a crystal will depend on direction and degree of work-hardening, obeying the resolved shear stress law (Schmid and Boas 1935). Schmid (1931) finds that for a cylindrical single crystal of any orientation the yield point gives a constant resolved stress on the active slip plane, equal to the critical stress.

When the slip plane is most suitably orientated the yield stress or applied stress required is a minimum. We have seen that the slip plane is most suitably orientated for slip during indentation when the hardness value is a minimum. Maximum hardness occurs when the crystal is least suitably orientated for slip or when the resolved shear stress on possible slip planes is least. Hardness of a crystal may therefore be considered to be a measure of the yield stress of the crystal in different directions. The property measured is the yield property mainly in a direction at right angles to the length of the indenter, though in fact it depends on the force distribution around the indenter which will determine the shape of the yield surface in the interior of the crystal. This agrees with the concept of the hardness of non-work-hardened material in a crystal as given by Tabor (1951).

ACKNOWLEDGMENT

Thanks are due to the Department of Scientific and Industrial Research for a Research Award which supported this work.

REFERENCES

BARRETT, C. S., 1943, *Structure of Metals* (New York : McGraw-Hill).
DANIELS, F. W., and DUNN, C. G., 1949, *Trans. Amer. Soc. Metals*, **41,** 419.
FRÉMONT, C., 1924, *Étud. Exp. de Technol. Industr. Paris*, **69,** 22.
GRODZINSKI, P., 1951, *Patent Spec.*, **65,** L 800, April 11 ; 1952, *Sheet Metal Ind.*, **29,** 908.
HOMER, E., and PLUMMER, H., 1938, *J. Inst. Met.*, **64,** 169.
MEINCKE, H., 1950, *Z. Metallk.*, **41,** 344 ; 1951, *Industr. Diam. Rev.*, **11,** 37.
MIDDLEBERG, G. A. G., 1886, *Engineering, Lond.*, **42,** 481.
O'NEILL, H., 1923, *J. Inst. Met.*, **30,** 299 ; 1934, *Hardness of Metals and its Measurement* (London : Chapman and Hall).
PERRYMAN, E. C. W., 1950, *Metal. Ind., Lond.*, **76,** 23.
PETERS, C., and KNOOP, F., 1940, *Metals & Alloys*, **13,** 292.
SCHMID, E., 1931, *Z. Elektrochem.*, **37,** 447.
SCHMID, E., and BOAS, W., 1935, *Kristallplastizität* (Berlin : Springer).
SCHULTZ, F., and HANEMANN, M., 1941, *Z. Metallk.*, **34,** 124.
TABOR, D., 1951, *Hardness of Metals* (Oxford : University Press).
TOLANSKY, S., 1948, *Multiple Beam Interferometry* (Oxford : Clarendon Press.)
TOLANSKY, S., and NICKOLS, D. G., 1949, *Nature, Lond.*, **164,** 840; 1952, *Phil. Mag.*, **43,** 410.

An Occasional Mode of Growth in Diamond

By S. TOLANSKY and SAYEDA H. EMARA
Royal Holloway College (University of London), Egham, Surrey

MS. received 19*th January* 1955, *and in amended form* 4*th April* 1955

Abstract. It is shown that in rare instances a mode of growth by plane sheets operates on the octahedral faces of diamond which leads to the formation of six-sided growth features containing alternate angles of approximately 90° and 150°. It is shown that the edges of these are effectively parallel to the directions ⟨431⟩. These may possibly arise through intersection of {221} with the (111) face.

§ 1. Introduction

In this laboratory an intensive systematic optical study is being made of growth characteristics on natural diamond faces. In this study several hundred octahedral faces have been subject to detailed examination by the microscope, by interferometry and by the light-profile microscope. It has been established that by far the great majority of these diamond faces grow by means of plane sheet growth which leads to trigon formation and to oppositely oriented triangular growth pyramids (for a recent discussion see Halperin (1954)). In both trigons and triangular growth pyramids the edges are strictly parallel to the three edges of the normally triangular face constituting the (111) natural octahedral face of the diamond.

In two exceptional instances we have found again a plane-sheet mode of growth, but this time the edges of the growth sheets are parallel to ⟨431⟩ and it is the object of this brief note to describe these relatively infrequent growth features which we have found on four separate faces. Three of these occur on different octahedral faces of one diamond and the fourth on another diamond.

§ 2. Observation

Case I.

The first example is remarkable for its approximate symmetry and gives us the clue to the others. It appears on one face of a good transparent octahedron from South Africa, of weight 0·63 carat. The feature is quite small indeed and is shown (× 1000) in figure 1 *a* (Plate). It consists of a hexagonal pattern and measurement reveals that the alternate angles are respectively approximately some 90° and 150°. The sides are almost, but not quite, equal in length. The angles are difficult to evaluate with any precision because the sides are not quite straight. The angles can be measured to within a little better than ½°. They are respectively, in degrees, 89½, 150, 92½, 144, 95, 150½. Their sum is 721½.

Examination by fringes of equal chromatic order (figure 1 *b*) (see Tolansky 1948) establishes that this feature is a plateau standing merely some 1000 Å high above a more extensive growth pyramid. It is certainly a growth feature for there is no evidence for the existence of any etch effects whatsoever on this diamond. This observed hexagonal pattern can be crystallographically described as having its growth edges more or less parallel to the directions ⟨431⟩. The calculated angles for a hexagon in these directions are respectively 87° 47′ and 152° 13′. The values in the observed figure are somewhat similar. Although there is only a crude parallelism to ⟨431⟩ the main fact is the differentiation from the general mass of growth features on octahedral faces, which consist of trigon or pyramidal growth features in which the three sides are, almost invariably, inclined to each other at angles very near to 60°.

The lines which form this hexagon could arise from intersection of (111) with any of many planes which include for example (221), (430), (103), (014), etc. Of these, we tend to conjecture in favour of (221) for the following reasons: It is the only set of planes which can produce simultaneously on each of all the (111) planes of the crystal both our hexagonal pattern and equilateral triangles, as in figure 2, and both patterns appear simultaneously on three faces of our crystal.

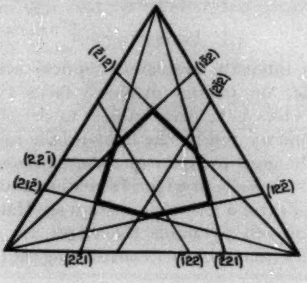


Figure 2.

Further, (221) is the most close packed of all the possible forms giving our hexagons and might well be favoured during growth. Again, it has been established by etching experiments (Omar, Pandya and Tolansky 1954) that the most important planes produced by etching (111) on diamond are (221). This evidence is far from decisive but favours (221) a little.

Case II.

In the second example, figure 1 *c* (×650), interferometry once more demonstrates that this figure lies on a growth pyramid. This pyramid reflects the feature, but there are distortions because of interference from surrounding growths.

As before, we have a six-sided figure, the angles being 90, 153, 92, 150, 92½, 151°. The sum of these is 728½° and the excess (8½°) over that for a rectilinear figure is undoubtedly due to curvature of the sides and not due to faulty estimation of the angles. Within the hexagon are the familiar 60° growth trigons.

Case III.

This example (figure 1 *d*) (×300) at first seems a most irregular pattern but can be classified along with the two previous ones. The key lies in the appearance of the alternate angles, approximating once more to some 90° and 150° which establish that this is just another distorted example. The angles measured are 89, 153, 93, 149, 90, 151, with sum 725°, and again the excess is attributable to curvatures.

It is noteworthy that these three features, each on a different face of one crystal, cover a relatively insignificant fraction of the total area. Despite the numerical differences in angles, these features are undoubtedly similar and are clearly related.

Case IV.

In this example (figure 1 *e*) (× 15), the mode of growth under consideration is on a much bigger scale and dominates the whole growth of the face. The stone is a clear octahedron mined in South Africa. There is a good deal of deviation from the ideal case shown in figure 1, nevertheless the relationship is evident. We have here a relatively rare example of a crystal face which has grown primarily both in this manner as well as by trigon formation. The strongly outlined hexagonal central feature, which seems to start off with angles near 90° and 150°, quickly transforms its character on moving to the edges, tending to become a crude 60° equilateral triangle.

We have sought through microphotographs published by other observers to see whether any examples of this particular hexagonal mode of growth have been recorded and overlooked. We have indeed found one clear example amongst 86 excellent micrographs of (111) faces recorded by A. F. Williams in his comprehensive book on diamond published in 1932. (There is amongst these photographs what may possibly be a second example but the evidence here is rather uncertain.) Although Williams had the opportunity of inspecting many thousands of stones yet he never, it seems, noticed this special mode of growth, despite its appearance in one of his published photographs. It is certainly of infrequent occurrence.

ACKNOWLEDGMENT

One of us (S. H. E.) wishes to express thanks to Ein Shams University, Cairo, for a grant of study leave enabling this work to be carried through.

REFERENCES

HALPERIN, A., 1954, *Proc. Phys. Soc.* B, **67**, 538.

OMAR, M., PANDYA, N. S., and TOLANSKY, S., 1954, *Proc. Roy. Soc.* A, **225**, 33.

TOLANSKY, S., 1948, *Multiple-beam Interferometry of Surfaces and Films* (Oxford : Clarendon Press).

WILLIAMS, A. F., 1932, *Genesis of Diamond* (London : Benn).

RESEARCH NOTES

Measurement of Minority Carrier Lifetimes in Semiconductors

By N. F. DURRANT
The Plessey Company Limited, Caswell, Towcester, Northants

MS. received 29th March 1955

The lifetime of minority carriers is an important property of semiconductors such as germanium and silicon and is one of the factors governing the selection of material for the manufacture of transistors and allied devices. Many methods for measuring this property have been described (Goucher 1951, Haynes and Shockley 1951, Navon *et al.* 1952, Pell 1953, Many 1954, Stevenson and Keyes 1955) but some of these have tended to give misleading results where trapping of minority carriers is an important effect (Arthur *et al.* 1955). The method to be described here is such as to allow trapping effects or conductivity modulation to be detected and steps taken to eliminate them.

A modification of the method for determining drift mobilities as described by Haynes and Westphal (1952) is used and their electrode configuration is retained. A short pulse of minority carriers, holes in the case of an n-type specimen, is injected into the specimen from an emitter. A sweeping field is applied to the specimen by a pulse which is triggered by, but delayed after, the emitter pulse. This delay may be varied at will. Under the influence of the sweeping field the carriers drift to the collector point. The holes decay after the interval owing to recombination so that the number passing the collector is proportional to $\exp(-t/\tau)$ where t is the period between injection and collection and τ is the lifetime. Diffusion of the carriers takes place according to the equation $L=(Dt)^{1/2}$ where L is the diffusion distance and D the diffusion constant. The collected pulse is not therefore a sharp replica of the emitter pulse but is rounded into a gaussian distribution whose area is proportional to $H\sqrt{t}$ where H is the collected pulse height. This area is proportional to the number of holes surviving so that $Ht^{1/2} \propto \exp(-t/\tau)$.

Two pulse generators are used in the arrangement shown in figure 1. The first, which may conveniently be free running at about 100 p.p.s. provides an

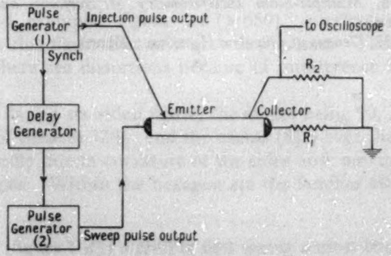


Figure 1. Experimental arrangement

emitter pulse of 2 μsec duration. The second generator, triggered from the first through a variable time delay circuit provides a sweep field of about 20 μsec duration. As can be seen from the figure, bias is applied to the collector during the pulse by the voltage drop across that part of the specimen to the right of the collector together with R_1, so that no d.c. bias need be applied. During the series of measurements, t is varied by altering the setting of the time delay generator without disturbing any of the other parameters such as injection efficiency or collector bias. Both t and H may be read from a calibrated oscilloscope (figure 2).

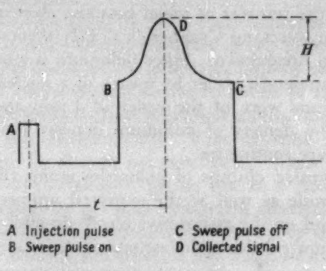


Figure 2. Typical oscilloscope display.

Conductivity modulation has the effect of delaying the arrival of the peak of the pulse so that the trailing edge of the collected pulse is made steeper. This effect may be reduced to negligible proportions by using a small emitter pulse. If trapping is present the trailing edge is extended by carriers released from traps. A uniform and constant illumination may be applied to the specimen so as to raise the overall concentration of minority carriers and saturate the trapping levels. This should also be applied if it is found that the efficiency of the collector is varying with carrier density (Jacobs *et al.* 1953, Hogarth 1954).

It is found that the above method may be conveniently and easily used to give results that are consistent and reproducible within an accuracy to about 5%.

ACKNOWLEDGMENTS

I am glad to acknowledge the encouragement and help of Mr. A. L. Gray and am indebted to the Plessey Company for permission to publish this note.

REFERENCES

ARTHUR, J. B., BARDSLEY, W., GIBSON, A. F., and HOGARTH, C. A., 1955, *Proc. Phys. Soc.* B, **68**, 121.
GOUCHER, F. S., 1951, *Phys. Rev.*, **81**, 475.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81**, 835.
HAYNES, J. R., and WESTPHAL, W. C., 1952, *Phys. Rev.*, **85**, 680.
HOGARTH, C. A., 1954, *Proc. Phys. Soc.* B, **67**, 636.
JACOBS, H., MATTHEI, W., and BRAND, F. A., 1953, *J. Appl. Phys.*, **24**, 1410.
MANY, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.
NAVON, D., BRAY, R., and FAN, H. Y., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 1342.
PELL, E. M., 1953, *Phys. Rev.*, **90**, 278.
STEVENSON, D. T., and KEYES, R. J., 1955, *J. Appl. Phys.*, **26**, 190.

The Transfer of Metal between Electrodes in a High Vacuum on Application of a High Electric Field

By P. F. BROWNE†

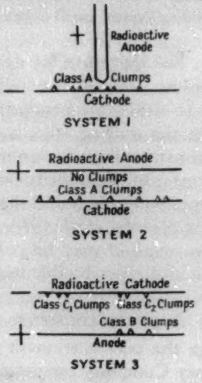
Physical Laboratory, Trinity College, University of Dublin

MS. received 30*th March* 1955

Deposits of polonium on nickel and lead were used as radioactive tracers to investigate the transfer of metal between electrodes in high vacuum with a view to checking Cranberg's (1952) theory of the initiation of high vacuum electrical breakdown. Since polonium is α-active, one can observe the distribution of transferred tracer by means of a nuclear plate.

Electrode separations were of the order of 1 mm and d.c. voltages up to 80 kv were used. The density of polonium deposit was usually of the order of 10^{11} atoms per square centimetre.

Nuclear plates revealed clumps of polonium atoms (figure 1, Plate) on the non-radioactive electrode as well as the expected uniform distribution. Such clumps were also found on the radioactive electrode after experiments in which it was the cathode. Figure 2 shows the various electrode systems and the different types of clumps which were obtained. Class A clumps were obtained at voltages under the sparking voltage as well as after sparking. All class B clumps were obtained after sparking. Class C clumps are divided into two groups, those obtained at voltages below the sparking voltage (class C_1) and those obtained after sparking (class C_2). Class C_2 clumps are much larger than class C_1 clumps and there is reason to suppose that each class C_2 clump has a corresponding class B one.



+
Radioactive Anode
Class A Clumps
−
Cathode
SYSTEM 1
Radioactive Anode
+
No Clumps
Class A Clumps
−
Cathode
SYSTEM 2
−
Radioactive Cathode
Class C₁ Clumps
Class C₂ Clumps
Class B Clumps
+
Anode
SYSTEM 3

Figure 2.

Clumps contained too many polonium atoms to be explained by aggregate recoil (Lawson 1919). Estimations of the size of the clumps by projecting α-particle tracks in a nuclear emulsion back to their origin were limited by measurement errors and by errors introduced because of α-particle scattering. They did show that the polonium atoms of a clump were confined, in most cases, to an area of less than 10^{-7} cm^2.

---

† Now at Department of Physics, University of Birmingham.

Class A clumps were investigated most thoroughly. A high field rather than a high voltage is effective in producing them. Occasional sparks at large electrode separations occurred without producing clumps. These sparks were probably what Arnal (1954) has termed ' microdischarges ', involving adsorbed gas atoms from the electrodes. It became progressively more difficult to obtain clumps with the use of the electrodes.

The only plausible explanation of these experiments is as follows. Class A clumps are formed by positive ions of polonium drawn to points of field emission. The positive ions are formed from the polonium deposit on the anode. Class C clumps also consist of polonium ions drawn to points of field emission, but the polonium ions are now formed from polonium atoms liberated from the same electrode by positive ion bombardment. This explains why class C_1 clumps are small and why class C_2 clumps are very large. Class B clumps are formed when the point of electron emission, giving rise to a class C_2 clump, melts or disrupts, and is transferred to the non-radioactive electrode. Continued use of the electrodes (conditioning) destroys many of the sharp irregularities necessary for field emission by melting and positive ion bombardment. This picture of what is happening between the electrodes prior to breakdown agrees very well with that which Bertein (1946) has described. Inter-electrode current pulses, numerous at first and gradually decreasing with duration of applied voltage, were detected by an oscilloscope at voltages below the sparking voltage.

There was no obvious difference, apart from magnitude, between clumps obtained before and after sparking. A spark, being usually defined by the emission of visible light, is an arbitrary term. A series resistance of the order of megohms imposed an upper limit to the intensity of sparks but many sparks were quenched naturally before they reached this intensity. Field emitted currents could be estimated from the number of transferred polonium atoms assuming a known density of deposit on the anode. Currents of 10^{-12} amp for 1 minute produced easily detectable clumps.

One hesitates to identify the clumps with those postulated by Cranberg for two reasons. Firstly class C clumps cannot be clumps of transferred electrode material. Secondly a clump can have a density of polonium atoms (minimum estimate) which is greater by two orders of magnitude than the density of polonium deposit on the radioactive electrode from which it originated. For many cases, however, this second reason is not valid, and there is no reason why such class A clumps should not be attributed to the transfer of a clump of electrode material. The distribution of transferred clumps in the following experiment is typical and particularly favours the Cranberg clump explanation. With an anode formed by the sharpened tip of a nickel wire of diameter 0·053 cm and a plane cathode, the electrode separation being about 0·026 cm, 19 clumps were found over an area of 1 cm^2 of cathode. If a clump is detached in the solid state by electrostatic forces the initial acceleration might be in any direction over a solid angle of almost 2π because of local field variation.

The electrostatic stress, $E^2/8\pi$, at the anode at breakdown was found to agree well with the tensile strength of the electrode. The breakdown field E was calculated by observing the onset voltage of field emission when the electrode was cathode. Breakdown occurred at 7·6 times this voltage for copper. Thus electrostatic forces are quite capable of detaching clumps near the breakdown voltage.

Using electrodes of radioactive cobalt and cadmium, it was shown that less than 2×10^{-9} g of electrode material was transferred prior to breakdown.

The uniformly scattered distribution of polonium atoms was transferred most easily from cathode to anode probably because of the tendency of polonium, which is electronegative, to form negative ions. The amount transferred from anode to cathode increased with the applied voltage and with the duration of the applied voltage and, except where field emission or a spark occurred, can be explained by aggregate recoil (Lawson 1919).

ACKNOWLEDGMENTS

I am indebted to Professor E. T. S. Walton and Professor J. H. J. Poole for suggesting the research and for valuable discussions during its execution.

REFERENCES

ARNAL, M. R., 1954, *C.R. Acad. Sci., Paris*, **238**, 2402, 2061.
BERTEIN, F., 1946, *C.R. Acad. Sci., Paris*, **222**, 63.
CRANBERG, L., 1952, *J. Appl. Phys.*, **23**, 518.
LAWSON, R. W., 1919, *Nature, Lond.*, **102**, 465.

Field Enhanced Cathodo-Luminescence in Magnesium Oxide

BY J. WOODS AND D. A. WRIGHT
Research Laboratories of The General Electric Co. Ltd., Wembley, Middx.

MS. received 16*th February* 1955 *and in final form* 28*th March* 1955

IT is well known (Jacobs 1951, Jacobs *et al.* 1952, Dobischek *et al.* 1953) that thin layers of magnesium oxide bombarded with electrons of a few hundred volts energy can under certain conditions give an effect similar to the 'Malter effect' (Malter 1936). That is to say the secondary emission coefficient is a function of the difference between the collector potential and the potential of the metal surface ('target') to which the magnesia is applied. The effect occurs because the magnesia surface charges up to a potential near that of the collector, so that most of the collector–target potential difference appears across the magnesia layer. The secondary emission coefficient rises with increase in this potential difference, and may reach values between 100 and 1000 before breakdown of the layer occurs. It is also well known that under electron bombardment magnesia exhibits a feeble cathodo-luminescence. The work described below was undertaken to determine whether the conditions leading to field enhancement of secondary emission would also lead to field enhancement of the luminescence, i.e. whether in a layer showing the Malter effect the light output was determined solely by the energy and current in the primary bombarding beam, or whether it varied with the collector–target potential difference.

Experiments have been carried out on three types of magnesia: (i) formed by evaporating magnesium through oxygen at 0·1 mm pressure, on to a pure nickel base, and then heating, (*a*) at 700°c in oxygen or at 900°c in a vacuum, (*b*) at 400°c in a vacuum, (ii) by burning magnesium in air and condensing on a nickel plate a thin layer of oxide smoke which was examined (*a*) after heating in a vacuum to 900°c before any bombardment was given and (*b*) after heating only to 400°c in a vacuum before bombardment, (iii) by suspending in a binder solution a fine magnesia powder (analar) obtained from a chemical supplier, and spraying to form a coating on nickel. This was baked in air at 500°c to decompose the binder, and was then examined after heating in a vacuum, again (*a*) at 900°c and (*b*) at 400°c.

The coatings were bombarded at 1–100 μA cm^{-2} using current from an oxide-coated cathode. The light was measured using an R.C.A. type 931A photomultiplier. All measurements were made in sealed-off gettered tubes following careful outgassing processes during pumping.

We refer firstly to the results on magnesia of types (i*a*), (ii*a*) and (iii*a*), all strongly heated in a vacuum before bombardment. All showed blue cathodo-luminescence which was maintained for many hours under all conditions of test, up to 1 mA cm^{-2} at 500 volts. With (i*a*) the light was examined spectrographically and a broad band was identified centred at 3·1 ev. With (ii*a*) the blue band was centred at 2·8 ev, while faint bands were present at 2·3, 2·04 and 1·95 ev. With all three types, rapid increase of secondary emission with potential difference between collector and target was observed, and with all three was found an associated increase in luminescence, without change in colour. At sufficiently high fields scintillations occurred as breakdown set in at different parts of the layer, but the luminescence increase was a genuine effect in no way connected with scintillations, occurring over a wide range of voltage below the breakdown threshold. This applied in particular to layers of types (i*a*) and (ii*a*); with the sprayed coatings (iii*a*) local breakdowns occurred at rather low mean fields, and there was not such a wide range in which luminescence enhancement occurred, nor was it possible to apply high mean fields, as they led to cumulative breakdown processes.

The diagrams show the increase in secondary emission current and light output occurring (figure 1) on keeping the target potential constant and raising the collector potential and (figure 2) on keeping the collector potential constant and *lowering* the target potential, the two potentials being initially equal in each case.

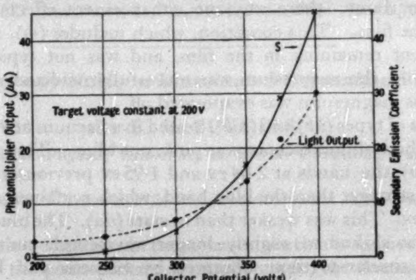


Figure 1. Secondary emission and light output as a function of collector voltage. Target voltage constant.

The genuine nature of the effect is perhaps demonstrated most clearly from the second result. A plot is also given (figure 3) of the variation of light with voltage when the two potentials are varied in step, so that having set one, the other is adjusted to make the secondary emission coefficient unity. This gives the true variation of light output with target voltage, and is seen to give a curve much less steep than in figures 1 and 2, in which the enhancement effect occurs.

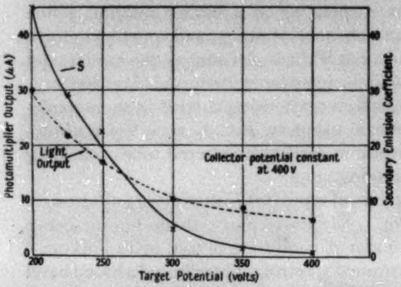


Figure 2. Secondary emission and light output as a function of target voltage. Collector voltage constant.

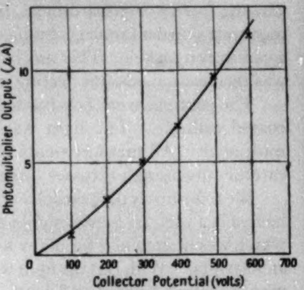


Figure 3. Light output as a function of collector voltage. Target voltage adjusted to maintain a secondary emission coefficient of unity.

The variation of secondary current with voltage follows the same type of law as investigated by Dobischek *et al.* (1953). It appears that electron multiplication processes are occurring in the oxide, leading to the increase in secondary current and in light. Thus the light increase is due to a d.c. electroluminescent effect. We find with layers (ii*a*) that having established the enhanced effect, the bombarding beam can be switched off, and a high current from target to collector is maintained for an hour or more. A high light output is maintained at the same time, and is now provided entirely by the high field phenomenon, with no cathodo-luminescence. Both current and light output are a little lower than when the bombarding beam is on.

After heating magnesia of type (i) at temperatures lower than those used for state (i*a*), i.e., lower than 700°c in oxygen or 900°c in a vacuum, the luminescence was blue but very faint; there were no enhancement effects because of high conductivity of the film. This condition, which includes (i*b*), was due to a high magnesium content remaining in the film, and was not typical of the oxide. In oxygen at 700°c the magnesium was more fully oxidized, and in a vacuum at 900°c the excess magnesium was evaporated off.

With magnesia of types (ii*b*) and (iii*b*), heated in a vacuum at 400°c before bombardment, the cathodo-luminescence was pink, not blue. This was because with both (ii*b*) and (iii*b*), the bands at 2·04 ev and 1·95 ev previously observed in (ii*a*) were now much stronger than the blue band, which remained stronger than the green band at 2·3 ev. This was weaker than in state (ii*a*). The blue was now centred at 2·7 ev i.e. it was shifted to slightly longer wavelength compared with (ii*a*). These bands are similar to those observed by Saksena and Pant (1954) after electron bombardment of single crystals of magnesia. Our plates were not sensitive beyond 6600Å, so that the red band at 1·73 ev (7100 Å) observed by Saksena and Pant may also have been present. The 1·95 ev band was considerably stronger than the 2·04 ev. The enhancement effect was present, but was smaller, both in secondary current and in light, than with (i*a*), (ii*a*) and (iii*a*). When once bombardment had occurred, e.g. at 100 μA cm^{-2} and 500 volts for 15 minutes, subsequent heating in a vacuum below 1100°c or in 0·1 mm of oxygen up to 800°c had no effect. The blue luminescent state with the large enhancement effect was not obtained, although when similar heating had been given before bombardment, with (i*a*) and (ii*a*), this was the stable state produced. Heating to between 1100 and 1200°c in a vacuum did, however, produce the blue luminescent state. This occurred also when smoke layers of type (ii*b*) deposited on tantalum or platinum were heated between 1100 and 1500°c in a vacuum. When bombardment was continued with types (ii*b*) and (iii*b*), there was rather rapid visible darkening of the magnesia, accompanied by a decrease in luminescence. Heating in a vacuum up to 1000°c then had no effect; heating in oxygen at 800°c removed the darkening, but the luminescence remained pink, and the enhancement remained small.

When the voltage across the layer was increased in the early stages of bombardment, there were violent current fluctuations and bright scintillations when breakdown occurred. The voltage difference leading to breakdown was similar in state (ii*b*) to that in (ii*a*). The magnesia remaining at regions where breakdown had produced partial local disruption of the layer, gave the blue luminescence and the more marked enhancement. This was probably due to the intense local heating during the breakdown process.

The nature of the red, green and blue emitting centres, the reason for the difference in scale of the enhancement effect in the two states, the mechanism of the enhancement and the reason for the remarkable effect in (ii*b*) and (iii*b*) of electron bombardment in stabilizing the red state, are the subjects of continued research.

There are indications that the blue band is due to excess magnesium, and the red to excess oxygen, but further evidence is required to establish definitely whether this is so.

References

Dobischek, D., Jacobs, H., and Freely, J., 1953, *Phys. Rev.*, **91**, 804.
Jacobs, H., Freely, J., and Brand, F. A., 1952, *Phys. Rev.*, **88**, 492.
Jacobs, H., 1951, *Phys. Rev.*, **84**, 877.
Malter, L., 1936, *Phys. Rev.*, **50**, 48.
Saksena, B. D., and Pant, L. M., 1954, *Proc. Phys. Soc.* B, **67**, 811.

Low Frequency Noise in Illuminated Germanium Diodes

By R. E. BURGESS
Department of Physics, University of British Columbia, Vancouver

MS. received 28*th December* 1954 *and in amended form* 28*th January* 1955

It is well known that when bias in the reverse direction is applied to a germanium diode (either of the point contact or junction type) the fluctuations of current have a spectral density which at low frequencies is greatly in excess of the expected thermal and shot noise components. This excess noise has a mean square current spectral density which varies approximately inversely as the frequency and increases with increasing current; at sufficiently high frequencies the excess component is negligible compared with the thermal and shot components.

The current flowing in such a diode under reverse bias may be increased by illumination and the current increment is fairly closely proportional to the light intensity. It is of fundamental importance for understanding the mechanism of noise generation in semiconductor devices to enquire how the noise varies under illumination; it is also of practical significance in assessing the sensitivity of such devices as photocells.

Experiments have been reported on low-frequency noise measurements of illuminated germanium diodes. Slocum and Shive (1954) made measurements of the noise in p–n junction diodes in a band centred on 1 kc/s; the diodes were selected for low reverse current and the noise measured at various bias voltages with varying illumination. It was found that for currents from about 5 to 200 μA the noise intensity was very close to the shot noise value. Earlier observations by Shive (1952) at 1 kc/s can be analysed to support the view that the *increase* of noise at a given reverse bias when the diode is illuminated corresponds to shot noise associated with the increase of current.

Granville and Gibson (1953) made measurements at 16·5 c/s on a reverse biased point-contact diode with and without illumination and noted that *at a given current* the noise was less under illumination than when dark. The authors interpreted this result qualitatively in terms of a reduction of the effective α in the current magnification factor for the hole to the contact. It will now be shown that a simpler interpretation will explain these results as well as those of Slocum and Shive, namely, that whereas the dark current may have an appreciable low-frequency component of noise which is unaffected by illumination, the increase of current due to illumination (at a given reverse bias) is accompanied by a corresonding addition of shot noise only.

The elementary theories (Lehovec 1948, Cummerow 1954) of the current–voltage characteristic of an ideal p–n junction or point-contact diode under irradiation yield

$$I = A[1-\exp(-\boldsymbol{e}V/\boldsymbol{k}T)] + BL$$

in which V is the applied voltage and I the diode current (both measured in the reverse direction), $\boldsymbol{e}$ the electronic charge, $\boldsymbol{k}$ Boltzmann's constant, T absolute temperature and L intensity of illumination; A and B are constants which depend on the properties of the diode and the spectral distribution of the incident light flux. In practical diodes A and B both increase with increasing reverse voltage and this non-ideal behaviour may be attributed to one or more of the following effects : carrier pairs generated within the depletion layer (Shockley and Read 1952), leakage across a surface barrier (Culter and Bath 1954) and electron multiplication in regions of high electric field (McKay and McAfee 1953).

Thus the theory for the ideal diode indicates that illumination produces an increase of reverse current which is independent of the applied bias voltage. From figure 1 of Granville and Gibson (1953), it is seen that the increment of reverse current was in fact nearly constant at 8 μA for the illumination used. Now it is to be expected that the dark current in these experiments at the low frequency of 16·5 c/s contained very great excess noise, i.e. greater than shot noise by a factor of at least 10^3. Consequently the addition of $8\,\mu$A of shot noise on illumination would result in a completely negligible increase of noise and this is in fact closely confirmed on a quantitative basis by the authors' figure 1 when comparison of noise *at a given reverse voltage* is made. It is suggested that their comparison at a given reverse current is quite inappropriate for noise mechanism interpretation since this involves different reverse voltages in the dark and illuminated conditions.

In the early experiments of Shive on the junction diode at 1 kc/s the excess noise of the dark current was not so overwhelming as in the point-contact diode and accordingly when sufficient illumination was applied so as to increase the dark current by a factor of 100 the increments of noise were detectable and numerical analysis shows them to be of the order expected for the shot noise of the photocurrent (i.e. $14\,\mu\mu$A noise in a 1 c/s bandwidth for $600\,\mu$A current).

In the later experiments of Slocum and Shive the junction diodes were of improved technique of manufacture and selected for low dark current. Accordingly the excess noise of the dark current was relatively small, and this resulted in a close approach to shot noise behaviour under all conditions of applied bias and illumination.

The fundamental significance of this interpretation of the experimental data is of importance in understanding the mechanism of excess noise generation at low frequencies in semiconducting devices. The dark reverse current of a diode arises from the hole–electron pairs generated within a few minority-carrier diffusion lengths of the barrier region; the additional photocurrent arises from the pairs generated by the incident light quanta within the same region. Since these processes are therefore alike except in the actual pair generation mechanism it is concluded that the dark-pair generation is subject to a modulation process with a marked low-frequency spectrum. Whereas the photoelectric generation of pairs arises from direct electron transitions between the valence and conduction bands the dark current generation process mainly arises (at least in germanium) by way of intermediate bound levels (Shockley and Read 1952). It therefore seems that the capture and release of electrons by these intermediate levels or traps does not proceed randomly and independently. Either co-operative effects may exist between traps which are close together or, more probably, the rate at which electron transitions proceed to and from individual traps is not a simple Poisson process.

REFERENCES

CUMMEROW, R. L., 1954, *Phys. Rev.*, **95**, 16.
CUTLER, M., and BATH, H. M., 1954, *J. Appl. Phys.*, **25**, 1440.
GRANVILLE, J. W., and GIBSON, A. F., 1953, *Proc. Phys. Soc.* B, **66**, 1118.
LEHOVEC, K., 1948, *Phys. Rev.*, **74**, 463.
MCKAY, K. G., and MCAFEE, K. B., 1953, *Phys. Rev.*, **91**, 1079.
SHIVE, J. N., 1952, *Proc. Inst. Radio Engrs*, **40**, 1410.
SHOCKLEY, W., and READ, W. T., 1952, *Phys. Rev.*, **87**, 835.
SLOCUM, A., and SHIVE, J. N., 1954, *J. Appl. Phys.*, **25**, 406.

REVIEWS OF BOOKS

Building Materials: their Elasticity and Anelasticity, edited by M. Reiner. Pp. xv + 560 (Amsterdam : North-Holland Publishing Company, 1954) 84*s*.

Books made up of contributions by a number of specialist authors are becoming increasingly common, and these are always difficult to review adequately. This arises because of the distinction which should be made between good articles and good collections of articles. In the present case the frame of reference set by the Editor is sound and worth while, the articles are to the point, and the whole work is clearly going to be of value in building research. In recent years, a lot of information has reached the technical press on the behaviour of such materials as clays, concrete, asphalt and wood, and a collation at this stage is timely. The main concern of the authors is with the rheological properties of materials, and in this connection, Professor Reiner contributes a well-assembled introductory article, with Professor Freudenthal continuing with the structural engineering aspects. The complexities of time- and rate-dependent processes are always difficult to present; the phenomena are here neatly sorted out and many will find these articles the most valuable in the book. They should, indeed, be recommended to anyone starting out on plasticity problems. For the rest, the book covers a whole range of materials; paints, putties, metals, soils, plasters and so forth. The style and the subject matter are more academic than practical, and although it would be ungracious not to welcome this enterprising and interesting work, it must be said that it will be the research worker rather than the applied engineer who will benefit directly from it. As the book was designed particularly for the technologist, it is a pity that the power of the methods is not brought out more specifically by reference to actual civil engineering problems and projects. It is only in this way that rheological ideas will gain wider acceptance in the applied field.

A. J. KENNEDY.

Table of the Gamma Function for Complex Arguments. National Bureau of Standards—Applied Mathematics Series 34. Pp. xvi + 105. (Washington : U.S. Government Printing Office, 1954.) $2.

Once again we are indebted to the Mathematical Tables Project of the National Bureau of Standards for a table of considerable importance to many engaged in computation in both pure and applied mathematics. In physical problems, the gamma function with complex argument often occurs as an external factor to complex hypergeometric series, and may need to be known with unusually high accuracy.

The table under review presents 12D values of the real and imaginary parts of $\log_e \Gamma(z)$ for $z = x + iy$, $x = 0(0{\cdot}1)10$, $y = 0(0{\cdot}1)10$. By means of the familiar $\sin \pi z$ formula, and the duplication formula, the values given can readily be extended to cover the ranges, $-18 \leqslant x \leqslant 20$, $-20 \leqslant y \leqslant 20$. Outside this region the asymptotic series of Stirling provides similar precision, so that in effect the quantitative behaviour of $\Gamma(z)$ in the whole complex plane is known. In a table of such wide range it is of course unreasonable to expect accurate interpolation to be everywhere straightforward. Linear interpolation is out of the question except for the crudest calculations, but one finds that a simple 4-point interpolation mostly provides 5D accuracy when $x, y > 1$, improving to 8D accuracy for larger values of x and y. For uses needing greater precision details are given of interpolation formulae using up to 8 points in the table, and ensuring at least 9D accuracy when $x, y > 2$. Various formulae are also suggested for inverse interpolation.

There remains of course the awkward region near the singular point at $z = 0$. Here the obvious procedure is to use this table to find $\Gamma(2+z)$ or $\Gamma(1+z)$, and hence $\Gamma(z) = \Gamma(1+z)/z$. There also exists the complementary table of Stanley and Wilkes (Toronto, 1950), giving 6D values of $1/\Gamma(z)$ for $x = -0{\cdot}5(0{\cdot}01)0{\cdot}5$, $y = 0(0{\cdot}01)1$, which remains a valuable accessory. Otherwise the N.B.S. table has no rival, and is good value at the price.

R. A. B.

Advances in Electronics and Electron Physics, Vol. VI, edited by L. Marton. Pp. xii + 538. (London and New York: Academic Books Ltd., 1954) $11.80.

With the latest volume the title of this well known series has been changed by adding "Electron Physics" to the original "Electronics." In fact the previous volumes have always been concerned with the wider field, containing only a minority of contributions devoted to the subject of Electronics in the narrow meaning which has come to be attached to the term. It is clearly the intention of the Editor to concentrate even more in future on the fundamental phenomena in which electrons take part. A survey of the topics treated, which in any case is all that can be attempted here, shows how wide he casts his net.

The first review is of Metallic Conduction at High Frequencies and Low Temperatures, by A. B. Pippard, which is primarily concerned with the anomalous skin effect at micro-wave frequencies and its importance in the study of superconductivity. A short article by E. Abrahams deals with Relaxation Processes in Ferromagnetism, under which are included the frequency dependence of initial permeability, ferromagnetic spin resonance absorption, heat transfer effects and lag of magnetization in a changing magnetic field. Two authors from the Philips Laboratories at Eindhoven, J. Smit and H. P. J. Wijn, write on the Physical Properties of Ferrites. They cover, in fact, the preparation of ferrites and their behaviour under operating conditions as well as their fundamental physical properties. A contribution of over 100 pages from H. F. Ivey provides a most valuable account of recent experimental and theoretical work on Space Charge Limited Currents, ending with a discussion of the related problems that arise in semiconductors.

Two other articles on electronics in the narrower sense are those by W. M. Webster, comparing Analogous Semiconductor and Gaseous Electronic Devices, and by R. G. E. Hutter on Travelling Wave Tubes. In each of these a large amount of recent development is surveyed. A subject which involves the electron both in its 'free' aspect and in its interaction with matter—the Electron Microscope—is dealt with in a comprehensive review by M. E. Haine. The final contribution is by J. van der Handel on Paramagnetism.

It will be seen that the Editor has given a very wide interpretation to Electron Physics, including problems of the solid state as well as the emission of electrons from solids and the behaviour and utilization of the free electron. Detailed criticism is hardly possible with such a catholic collection, but it should be said that the reviewer found admirable each of the three surveys on which he felt able to pass judgment. Dr. Marton is to be congratulated on the feat of maintaining a high standard of contributions into the sixth volume of the series, with promise of more to come.

V. E. COSSLETT.

Applied Mass Spectrometry. The Report of a Conference organized by the Mass Spectrometry Panel of the Institute of Petroleum, held in London, October, 1953. Pp. vii + 333. (London : Institute of Petroleum, 1954). 50*s*.

As the title of this Report would suggest, the bulk of the papers contributed to this Conference deal with applications of mass spectrometry, mainly to problems of chemical analysis, but also to certain fundamental problems of chemistry and molecular physics. It is significant that comparatively little interest was shown (38 pages of the report) in the fundamentals and developments in ion optics; at least an equal interest was shown, for example, in computing methods in analysis. The conference held at the National Bureau of Standards in Washington some years ago included a much more comprehensive series of papers on the physics of mass spectrometers and it is a pity that this report is not more readily available in this country.

Atomic and molecular physicists will find the section on molecular structure of interest, in particular the papers contributed by G. A. MacDowell and other members of Liverpool University. There is also an excellent bibliography of the subject, covering the period September 1950—June 1953.

J. B. HASTED.

Microscope Design and Construction, by B. O. PAYNE. Pp. 271. (York: Cooke, Troughton & Simms Ltd. 1954.) 25*s*.

Intended for the scientific worker who requires to use a microscope as an implement in his work, this book competently describes both the optical and mechanical parts of modern microscopes. Also included is a description of illumination techniques and light sources. It is, of course, inevitable in such a book that the examples are primarily drawn from the range manufactured by the author's firm : special products of other companies have, however, not been excluded. It is particularly welcome to see a section on interference microscopy included with the phase-contrast technique. Detailed points of design, such as the statement that Lister type objectives of numerical aperture 0·3 may be freed from zonal spherical aberration and the implication that the bell crank lever slow motion may be expected to have as much as five microns of lost motion, are not above reproach, but the book is a careful piece of work and the comprehensive bibliography at the end of each chapter is of great value to the reader who wishes to study the subject matter of the chapter in greater detail.

The author is to be congratulated on producing an interesting book which is eminently suitable for its purpose, which is to assist the scientific worker to select equipment of a suitable type, and to make the best use of his apparatus.

G. W. HAMSTEAD.

Éléments de base des mesures en vol, by ROGER BETEILLE. Pp. 247. (Paris: Publication Scientifiques et Techniques de l'Air, 1954.) 2,230*fr*.

R. Beteille, who is a French aeronautical engineer and a specialist in the instrumentation for the flight testing of aircraft, has considered the dynamical theory of linear instruments and its application to the recording of pressure and temperature in flight, with particular attention to the derivation of height, Mach number, temperature and speed.

The interpretation of instrument readings under rapidly varying conditions is frequently necessary, as in the testing of fighter aircraft through the sound barrier, and even in the case of high performance transport aircraft, such as the Comet, where height may be varying continuously over a considerable part of the journey.

In part 1, Beteille discusses the conditions under which the input and output of an instrument can be related by a linear differential equation and uses the normal techniques of Fourier or Laplace transformation to obtain transfer functions, weighting functions, etc., giving a detailed treatment of the first and second order equations with constant coefficients.

Part 2 concerns the measurement of pressure, and he shows for example that the variation of pressure in an altimeter capsule, etc., connected by piping to a pressure head, can be approximately represented by a first-order equation with a characteristic pneumatic delay time varying as the ratio of the viscosity to the pressure. An acoustic delay time is also introduced.

Part 3 deals in a similar manner with the response times of the thermometers used in the measurement of surface or air temperatures, and with such topics as frictional and compressional heating, and the precautions necessary to avoid errors arising from radiation or conduction losses.

The last section treats the measurement of Mach number in subsonic and supersonic flight, and the measurement of height. A useful bibliography is appended, but there are few references in the text itself.

Because of the wide range, the treatment of several topics is necessarily superficial, and the book would probably be more rewarding if it were not biased quite so heavily toward a formal mathematical treatment, and dealt in greater detail with the instrumental layout, and the conduct of tests.

R. W. H. STEVENSON.

CONTENTS OF SECTION A

PAGE

Mr. W. HAYES and Prof. T. E. NEVIN. New Band Systems of MnF and MnBr in the Red 665

Mr. W. HAYES. The Spectrum of MnBr in the Photographic Infra-red . . 670

Dr. J. M. RADCLIFFE. On Lattice Scattering in Homopolar Semiconductors. . 675

Dr. R. A. ALLEN, Prof. W. E. BURCHAM, Dr. K. F. CHACKETT, Mr. G. L. MUNDAY and Dr. P. REASBECK. Electron Capture in the Decay of Sodium 22 . . 681

Dr. B. SCHIFF. The Cellular Method for a Close-packed Hexagonal Lattice, with applications to Titanium. 686

Dr. T. K. ALLEN. A Spectroscopic Study of Plasma–Electron Oscillations . . 696

Mr. H. R. LEMMER, Mr. O. J. A. SEGAERT and Dr. M. A. GRACE. The Decay of Cobalt 57 701

Mr. C. R. STANLEY. A New Phenomenon associated with Active Nitrogen . . 709

Dr. H. R. MOORE and the late Prof. F. C. CHALKLIN. The K-Series Emission Spectrum of Neon. 717

Dr. N. H. MARCH. Equations of State of Elements from the Thomas–Fermi Theory 726

Dr. G. M. LEWIS. The Decay of RaD and the 46·5 kev Excited State of RaE . 735

Dr. L. R. B. ELTON. Effect of Electron Mass on High Energy Electron Scattering 741

Dr. S. HOCHBERG, Prof. H. S. W. MASSEY, Dr. H. ROBERTSON and Mr. L. H. UNDERHILL. The Scattering of Nucleons by Alpha Particles—Determination of the Spin–Orbit Interaction 746

Dr. F. A. EL BEDEWI and Dr. M. A. EL WAHAB. An Investigation of the Neutron Groups from the Disintegration of Phosphorus by 8 MEV Deuterons . . 754

Research Note :

Dr. A. E. SOUCH. Decay of ^{70}As Produced by ^{14}N Bombardment of Cu. . 760

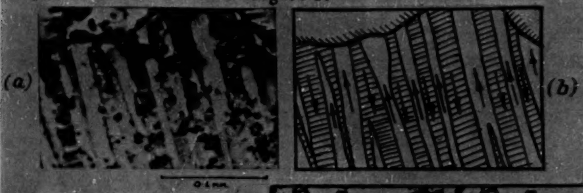
Letter to the Editor :

Dr. D. H. PARKINSON and Mr. J. E. QUARRINGTON. The Atomic Heat of Sodium 762

Reviews of Books. 764

Contents of Section B 768

Figure 1.



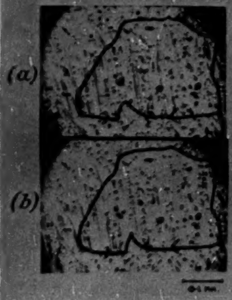


Figure 2.

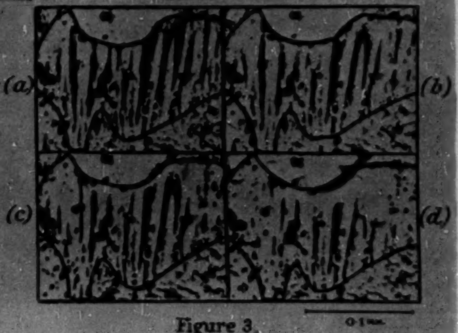


Figure 3.

Figure 1. (*a*) Powder patterns on a crystallite of specimen S1 (coercivity 2 Oe). (*b*) Diagrammatic interpretation of (*a*).

Figure 2. (*a*) Powder patterns on a crystallite of specimen S2 with a small normal applied field (coercivity 10 Oe). (*b*) Patterns on the same crystallite with the normal field reversed. Note the change of the powder deposits from one set of alternate domains to the other.

Figure 3. (*a*)–(*d*) Patterns showing the change in domain structure of a crystallite of specimen S1 when an applied field parallel to domain length is increased from zero.

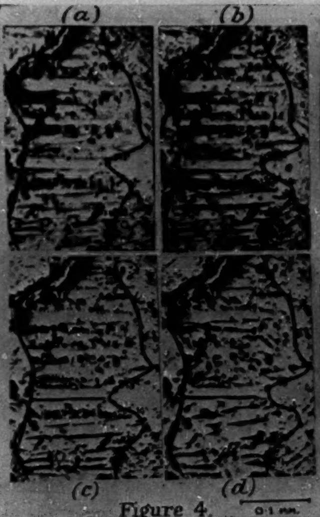


Figure 4.

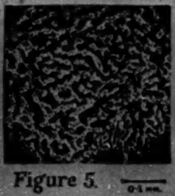


Figure 5.

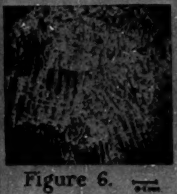


Figure 6.

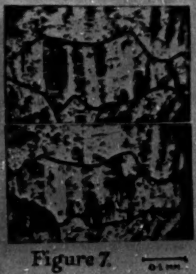


Figure 7.

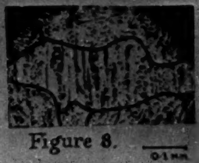


Figure 8.

Figure 4. (*a*)–(*d*) Patterns showing the change in domain structure of a crystallite of specimen S1 when an applied field perpendicular to domain length is increased.

Figure 5. Maze or honeycomb pattern on a crystallite of specimen S3 (coercivity 40 Oe).

Figure 6. A number or crystallites, in specimen S2, of different orientations showing several types of pattern.

Figure 7. (*a*), (*b*) Patterns obtained on a crystallite of specimen S3 after thermal demagnetization : (*a*) with a small normal field, (*b*) with the normal field reversed. Note that the deposits ‘ change domains ’.

Figure 8. The same crystallite as in figure 7 but after demagnetization by field reversals. The domains have decreased in size.

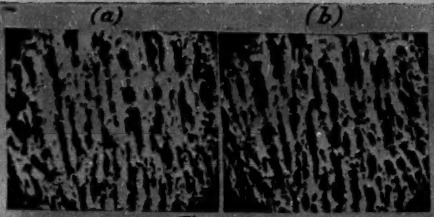


Figure 9.

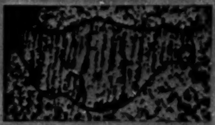


Figure 11.

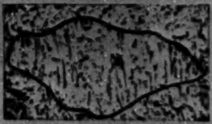


Figure 12.

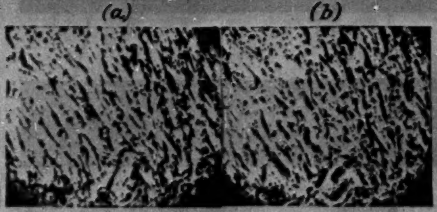


Figure 10.

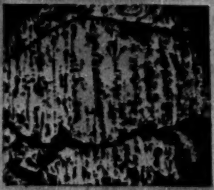


Figure 13.

Figures 9 and 10. Patterns obtained on a crystallite of specimen S4 (coercivity 100 Oe): (*a*) with a small normal applied field, (*b*) with the normal field reversed. The deposits have 'changed domains'.

Figures 11, 12 and 13. Patterns obtained on crystallites of specimen S4.

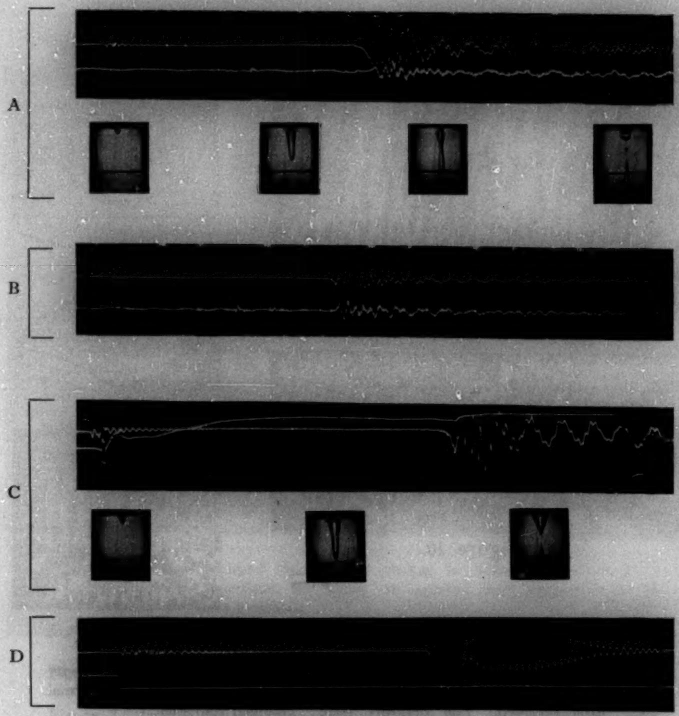
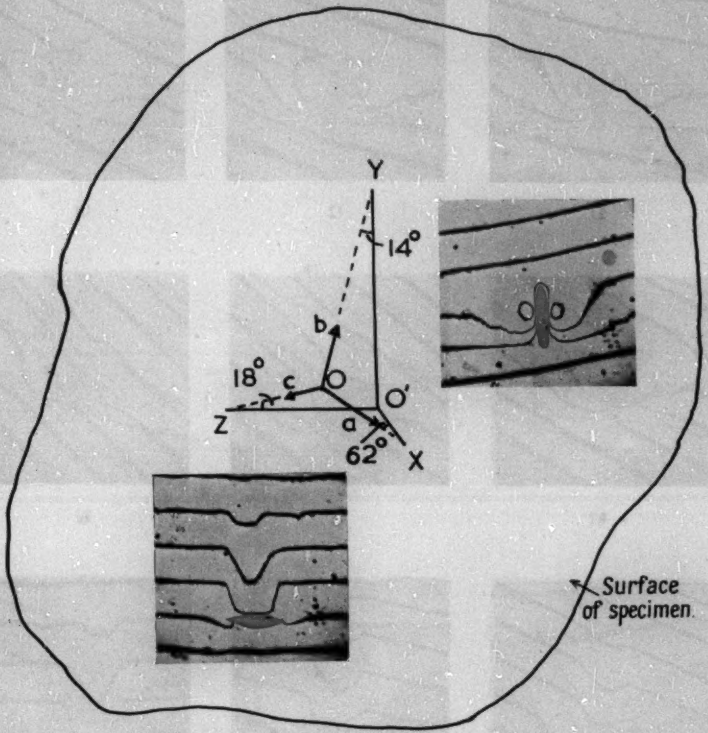


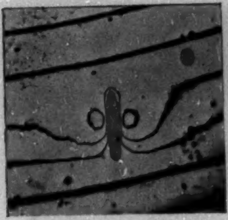
Figure 4. Sound records of cavity vibrations; A, B, hydrophone (above) and microphone (below): C, hydrophone and contact; these are for the 3 in. aluminium sphere dropped on water from 6 ft with photographs of the cavity at corresponding instants annexed; D, hydrophone and contact, 2 in. wooden sphere from 2 ft.

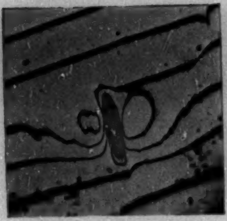

Y
14°
b
18°
c
O
O′
Z
a
62°
X
Surface of specimen

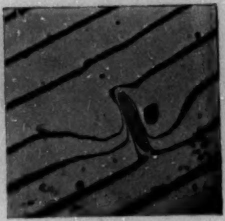
OX, OY, OZ axial directions, O′X, O′Y, O′Z projections of their directions on the crystal plane, which is in the plane of the paper.

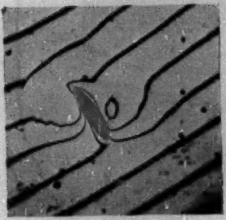
PLATE I.

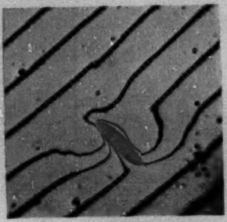
PROC. PHYS. SOC. VOL. 68, PT. 8—B (S. TOLANSKY AND A. P. WILLIAMS)

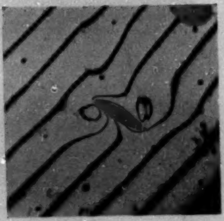

1

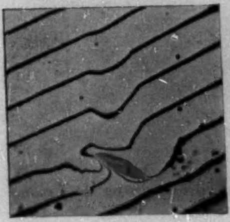

2

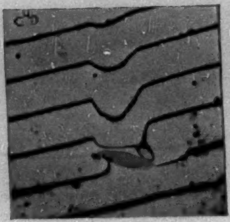

3


4


5


6


7


8

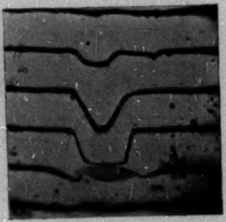
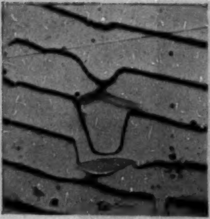
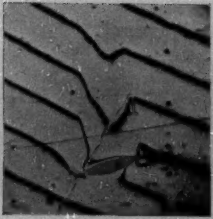

9

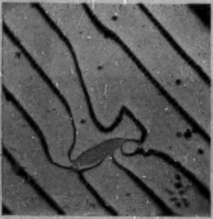
PLATE II.



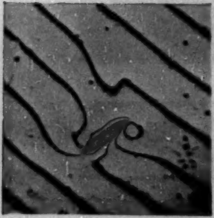
10



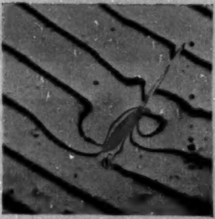
11



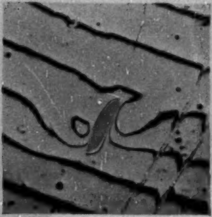
12



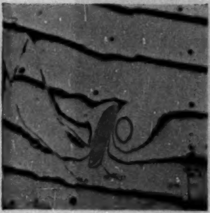
13



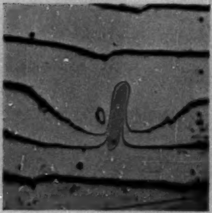
14



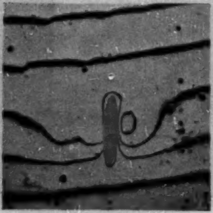
15



16

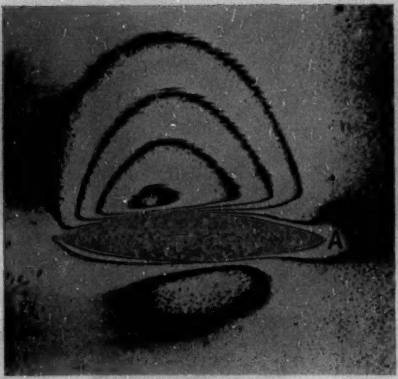


17

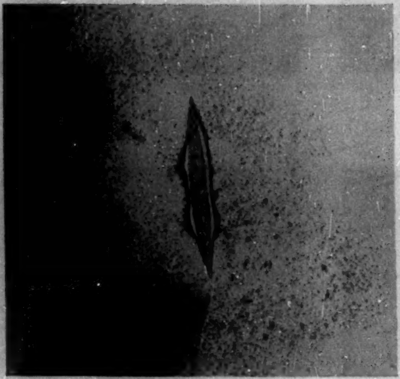


18

PLATE III.



1



2

PLATE IV.

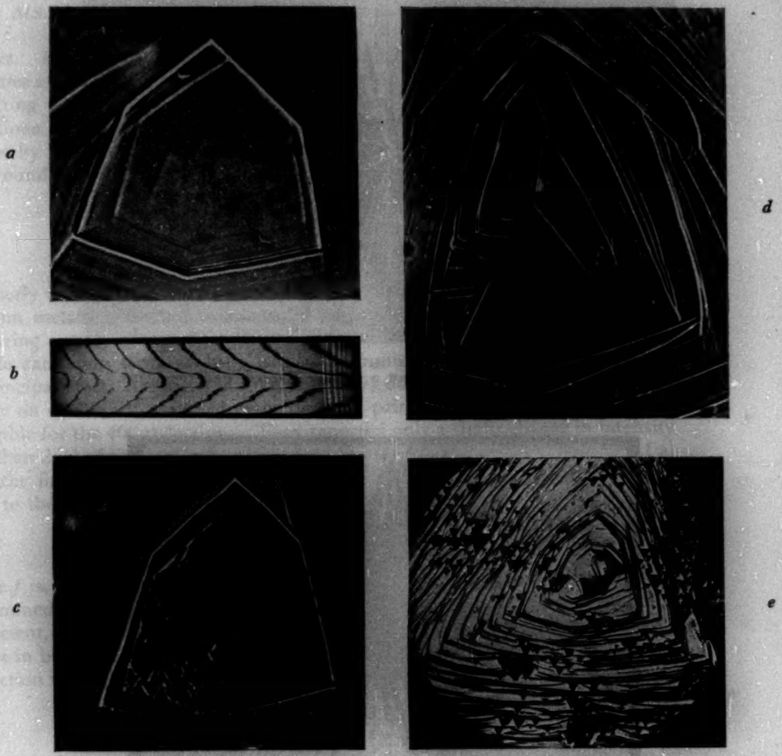
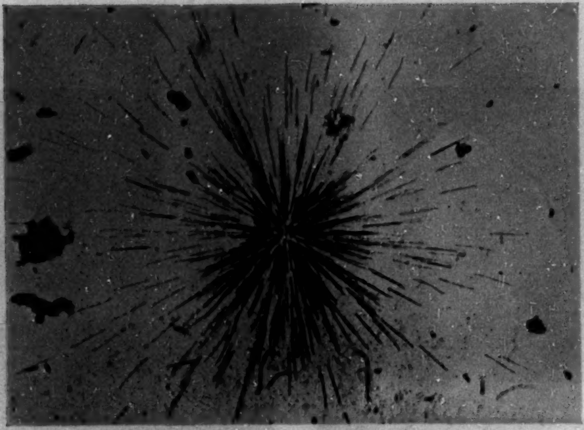
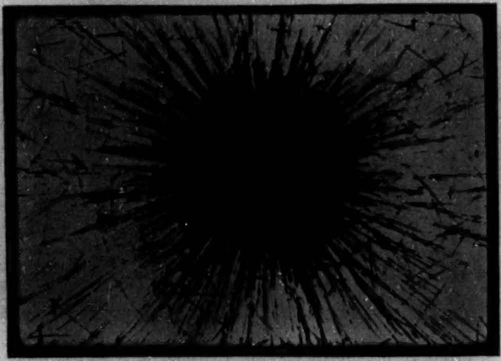


Figure 1. *a*, ×1000; *b*, ×50; *c*, ×650; *d*, ×300; *e*, ×15.



Class A clump obtained without a spark. Exposure of two days.



Class A clump obtained after a spark. Exposure of about one day.
Magnification—longest track is $2{\cdot}1 \times 10^{-3}$ cm.

Figure 1.

X-Ray Extinction and the Effect of Cold Work on Integrated Intensities

BY G. K. WILLIAMSON AND R. E. SMALLMAN
Metallurgy Division, Atomic Energy Research Establishment, Harwell, Berks.

MS. received 10th November 1954 and in amended form 27th April 1955

Abstract. Previous results on the intrinsic effect of cold work on integrated intensities are interpreted using the latest extinction corrections and a new method of plotting which gives reasonable differentiation between the various possible corrections. It is shown that cold work reduces the integrated intensities of the line by a small factor of between 3 and 5% and this intensity goes into the background as overlapping long tails to the lines.

§ 1. INTRODUCTION

In early studies of the effect of cold work on the integrated X-ray intensities from metals, extinction was usually neglected. However, Brindley (1940) during a study of the effect of cold work on the X-ray diffraction pattern did include extinction terms in his analysis, assuming that primary extinction had the principal effect. Averbach and Warren (1949), during their study of the intensity change on cold work, assumed extinction was primary (equation (2)) and hence negligible for the 400 reflection. They reported only a slight change in intensity (less than 2%) of this reflection upon cold work. Hall and Williamson (1951 a) were the first to investigate the experimental validity of extinction corrections prior to their study of cold work effects and proposed a correction of the form

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = 1 - gf^2 \qquad \ldots\ldots(1)$$

where f is the value of the atomic scattering factor, f_{obs} is their experimentally determined value uncorrected for extinction and g is the required extinction coefficient, an angular independent constant. This correction gave corrected f values in better agreement with theory than the Darwin correction for primary extinction which is of the form

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = \frac{\tanh pq}{pq} \simeq 1 - \tfrac{1}{3}(pq)^2 \qquad \ldots\ldots(2)$$

where p is the number of planes over which the lattice is perfect and q is given by

$$q = \frac{Na\lambda e^2}{\sin\theta mc^2} |F| \times \text{polarization factor}$$

where N is the number of unit cells per cm^3, a is the unit cell size, F is the structure factor and θ is the Bragg angle.

Weiss (1952 a) has quite correctly criticized the use of both of these corrections, and has suggested the following alternatives, for primary extinction:

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = 1 - g'f^2K^2 \qquad \ldots\ldots(3)$$

and for secondary extinction:

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = 1 - g'' \frac{f^2}{\sin\theta} K \qquad \ldots\ldots(4)$$

where K is the polarization factor, which, for crystal monochromatized radiation (Bragg angle α), takes the form (Guinier 1951):

$$K = \frac{1+\cos^2 2\alpha \cos^2 2\theta}{1+\cos^2 2\alpha} \qquad \ldots\ldots(5)$$

and g' and g'' are quantities independent of θ and depending primarily on the dimensions and the angular distribution of the mosaic blocks.

These corrections ((3) and (4)) have been criticized by Lang† (1953) who prefers corrections of the form given in (6) and (7) for primary and secondary extinction respectively:

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = 1 - g''' f^2 K \qquad \ldots\ldots(6)$$

$$\left(\frac{f_{\text{obs}}}{f}\right)^2 = 1 - g^{\text{iv}} \frac{f^2}{\sin 2\theta} K. \qquad \ldots\ldots(7)$$

Lang gives a refined treatment of polarization effects and has shown that, for a beam monochromatized by reflection from a crystal with Bragg angle α, the appropriate value of the polarization term K to be used in equations (6) and (7) is K_{L} where

$$K_{\text{L}} = \frac{1+\cos^2 2\alpha \cos^4 2\theta}{1+\cos^2 2\alpha \cos^2 2\theta}. \qquad \ldots\ldots(8)$$

Thus, since neither the correction used by Hall and Williamson (equation (1)) nor the correction implicitly assumed by Averbach and Warren (equation (2)) have theoretical justification it is of importance to reconsider their interpretations.

In this respect it is of interest to note that K_{L} is almost independent of θ over the range of θ normally encountered. For the results of Hall and Williamson K_{L} varies by less than ±6% over the range studied and by less than ±2% over the ranges in which the strong lines 3, 4, 19, 20 and 24 occur. To a first approximation then equation (6) reduces to equation (1) which Hall and Williamson observed to be a good experimental fit. Their conclusions, which in fact depend on the use of this correction, are unlikely to be basically altered. None of the recent corrections closely resemble the Darwin primary correction (equation (2)) and hence interpretations based on it need careful re-examination.

§ 2. INTERPRETATION OF EXISTING EXPERIMENTAL RESULTS

In previous papers (Hall and Williamson 1951 a, b, Weiss 1952 a, Lang 1953) both the nature and the magnitude of extinction effects were assessed by fitting the experimental points to a set of curves calculated for a wide range of g values and showing the variation of extinction with $N(=h^2+k^2+l^2)$. This procedure, apart from involving unnecessary calculation, is unsatisfactory for differentiating between the various corrections. The measurements are not absolute and hence after a first approximate fit has been obtained the results require normalizing and refitting. If extinction is large, adjacent curves are very similar, so that a precise value of g is difficult to obtain. An added disadvantage for the present study is the difficulty of deducing reliable accuracy brackets for the corrected intensities.

---

† These criticisms were advanced independently by the authors, see Smallman (1953).

Greater discrimination can be obtained if all corrections are written in the form

$$\frac{\text{Relative intensity observed}}{\text{Calculated intensity}} = \left(\frac{f_{\text{obs}}}{f}\right)^2 = C[1-g(A(\theta)f^2)] \quad \ldots\ldots(9)$$

where g and $A(\theta)$ are the appropriate extinction coefficient and geometrical factor (including K) respectively, of equations (3), (4), (6) or (7); C is the constant necessary to place the observations on an absolute basis such that, if extinction is zero,

$$(f_{\text{obs}})^2 = Cf^2. \quad \ldots\ldots(10)$$

Thus a plot of $(f_{\text{obs}}/f)^2$ against $A(\theta)f^2$ will give a straight line with intercept C and slope Cg. The value of g involving the use of an extrapolated C value is very sensitive to slight experimental error particularly with large extinction.

Equation (10), however, also holds for the corrected intensities (since for these intensities extinction is zero) and hence if cold work does produce a uniform fractional change in the apparent scattering factor, by dividing equation (10) for the corrected f^2 values for the annealed specimen by the corresponding equation for the cold worked specimen

$$\frac{(f_{\text{annealed}})^2}{(f_{\text{cold worked}})^2} = \frac{C_{\text{ann}}}{C_{\text{cw}}}. \quad \ldots\ldots(11)$$

If there is no change in the apparent scattering factor on cold working, $C_{\text{ann}} = C_{\text{cw}}$, on the other hand if there is a reduction, $C_{\text{ann}}/C_{\text{cw}} > 1$.

In the case of α-brass the intensities from the cold worked sample show negligible extinction. This simplifies analysis since I_{cw} can be equated to I_{calc} in equation (9), so that the required ratio $C_{\text{ann}}/C_{\text{cw}}$, can be obtained directly from one graph. In such cases, therefore, an intercept of unity indicates no change in the scattering factor due to cold work. A reduction in the apparent scattering factor on cold work is indicated by an intercept greater than one by an amount equal to the square of the fractional change in the scattering factor.

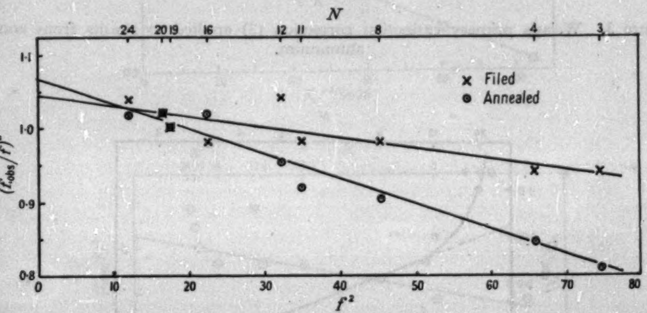


Figure 1. Application of correction (1) to results from annealed and filed aluminium.

This method of plotting has been employed on the results for aluminium (Hall and Williamson 1951 b) for all six corrections, figures 1 to 6. The results for aluminium are better for testing the validity of the various corrections than those for α-brass, because they include results from three groups of strong

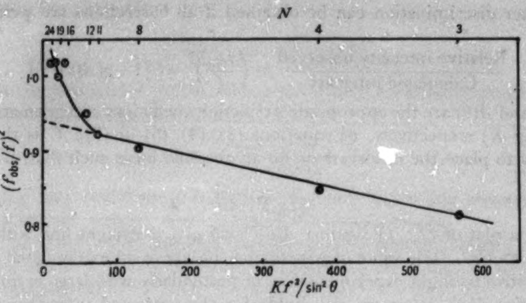


Figure 2. Application of the Darwin formula, correction (2) to results from annealed aluminium.

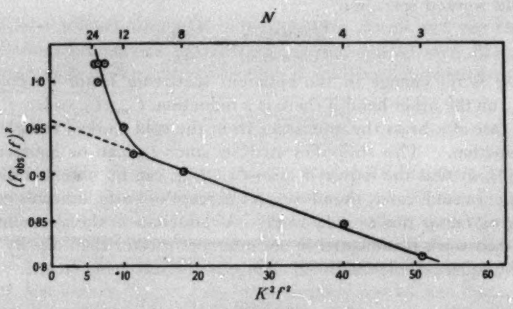


Figure 3. Weiss's primary extinction correction (3) applied to results from annealed aluminium.

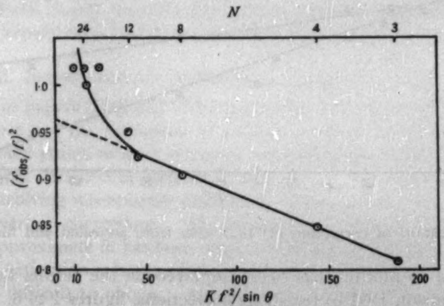


Figure 4. Weiss's secondary extinction correction (4) applied to results from annealed aluminium.

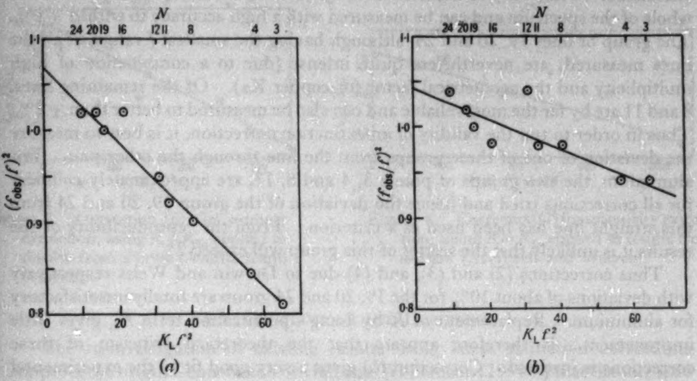


Figure 5. Application of correction (6), using $K=K_L$, to results from (*a*) annealed and (*b*) filed aluminium.

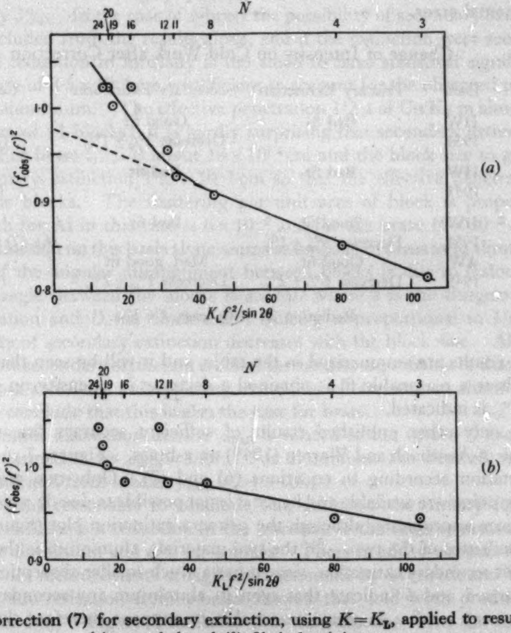


Figure 6. Correction (7) for secondary extinction, using $K=K_L$, applied to results from (*a*) annealed and (*b*) filed aluminium.

reflections. Thus the two adjacent lines 3 and 4 are the most intense pair in the whole of the spectrum and can be measured with a high accuracy to within ±1%. The group of lines 19, 20 and 24, although having the smallest f values of all the lines measured, are nevertheless quite intense (due to a combination of high multiplicity and the geometrical factor for copper Kα). Of the remaining lines, 8 and 11 are by far the most reliable and can also be measured to better than ±2%. Thus in order to test the validity of an extinction correction, it is best to measure the deviation of one of these groups from the line through the other pair. For aluminium, the two groups of points 3, 4 and 8, 11, are approximately collinear for all corrections tried and hence the deviation of the group 19, 20 and 24 from this straight line has been used as a criterion. From the reproducibility of the results it is unlikely that the scatter of this group will exceed 2%.

Thus corrections (2) and (3), and (4) due to Darwin and Weiss respectively with deviations of about 10% for the 19, 20 and 24 group are totally unsatisfactory for aluminium. Replacement of K by Lang's polarization term K_L gives little improvement. It therefore appears that the theoretical criticism of these corrections is justified. Correction (6) gives a very good fit to the experimental points, the deviations for any one group being less than 1%. This correction is slightly better than correction (1). Correction (7), however, (for secondary extinction) is not a good fit, the deviation being about 6%, outside probable experimental error.

Change of Intensity on Cold Work after Correction

Metal	Data	Primary Extinction	Secondary Extinction	Correction
Al	HW 1951	Bad fit	Good fit Decrease 3% ± 1%	Eqns (2), (1)
Al	HW 1951	Bad fit	Bad fit	Eqns (3), (4)
Al	HW 1951	Good fit Decrease 3% ± 1%	Bad fit	Eqns (6), (7) using $K=K_L$
α-Brass	AW 1949	Good fit Decrease 6% ± 2%	Med. good fit Decrease 1% ± 2%	Eqns (6), (7)

Radiation in each case Cu Kα.

The results are summarized in the table, and it will be seen that for the two cases where a reasonable fit is obtained a decrease in intensity on cold work of about 3% is indicated.

The only other published results of sufficient accuracy for interpretation are those of Averbach and Warren (1949) on α-brass. Figures 7 and 8 show the interpretation according to equations (6) and (7). Only two groups of high accuracy points are available and hence it is not possible to decide which correction is the more appropriate, although the primary extinction plot (figure 7) appears to be the better of the two. Of the two materials, aluminium is the more likely to exhibit secondary extinction, since it has a much smaller absorption coefficient, but figures 5 and 6 indicate that even in aluminium any secondary extinction effect is very small. It seems reasonable to suppose therefore that secondary extinction is most unlikely in α-brass and hence that the reduction in intensity on cold working is 6% ± 2%.

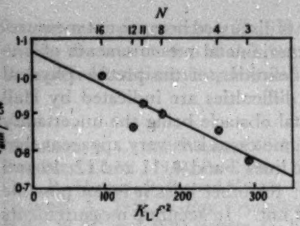


Figure 7. Correction (6) for primary extinction, using $K = K_L$, applied to results from α-brass (Averbach and Warren).

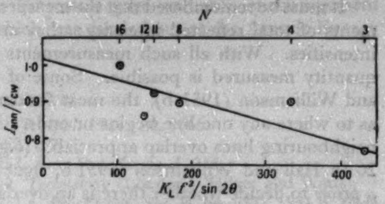


Figure 8. Correction (7) for secondary extinction, using $K = K_L$, applied to results from α-brass (Averbach and Warren).

§ 3. Conclusions and Discussion

The re-investigation of existing results using latest extinction corrections supports the original conclusion of Hall and Williamson (1951 b) that the intrinsic effect of cold work is to reduce the integrated intensities of the lines rather than the conclusion of Averbach and Warren (1949) that there is no change. In the case of aluminium extinction is largely primary and the reduction in intensity is approximately 3%. In the case of α-brass the possibility of secondary extinction cannot be excluded from the results alone, and if the extinction were secondary the indicated reduction in intensity is too small to have statistical significance.

A block size of $3{\cdot}5 \times 10^{-4}$ cm is sufficient to account for the observed primary extinction in aluminium. The effective penetration $1/2\mu$ of Cu Kα in aluminium is 40×10^{-4} cm or 11 blocks; it is hardly surprising that secondary extinction is negligible. For brass $1/2\mu$ is about 10×10^{-4} cm and the block size to give the observed primary extinction $1{\cdot}25 \times 10^{-4}$ cm so that the effective penetration is only 8 mosaic blocks. The scattering per unit area of block is proportional to Z^2D, which for Al in this case is 6×10^{-2} and for the brass 10×10^{-2} so that secondary extinction on this basis alone seems as unlikely in brass as in aluminium. In addition if the angular misalignment between blocks is due to dislocations, the smallest angle between the blocks is $\alpha = b/D$ where b is the Burgers vector of the dislocation and D the block size. Since g is proportional to $1/\alpha = D/b$ the probability of secondary extinction decreases with the block size. Allowing for these factors secondary extinction is about three times as probable in aluminium as it is in brass and since only primary extinction is observed in aluminium it seems safe to conclude that this is also the case for brass.

The minimum dislocation density $\rho_{\min}$ is related to the values D found for the primary extinction by $\rho_{\min} = 3/D^2$. For aluminium the derived value of $\rho_{\min}$ is $2{\cdot}4 \times 10^7$ and from α-brass $\rho_{\min} = 2{\cdot}4 \times 10^8$.

Thus it seems reasonable to conclude that extinction is primary for these two metals and there is a reduction in the intensity of the lines as measured by about 3% for aluminium and 6% for α-brass. There is added support for this latter figure since measurements of high precision made by Averbach and Warren for the weak 400 reflection from α-brass showed that cold work produced no change (within 2%). Figure 8 indicates that the reflection from the annealed sample was reduced by about 5% by extinction. Thus after correction the cold worked intensity is about 5% lower than the annealed one.

It must be remembered that the measurements discussed here are not measurements of total reflected intensity; they are experimental measurements of line intensities. With all such measurements no definition of the precise physical quantity measured is possible. Some of the difficulties are indicated by Hall and Williamson (1951 b), the most fundamental obstacle being the uncertainty as to where any one line begins or ends. In some cases it is very apparent that neighbouring lines overlap appreciably (e.g. see lines 3 and 4, 11 and 12, 19 and 20 in Hall and Williamson (1951 b, figure 1); for other lines it is not possible *a priori* to decide whether there is an overlap or not. In accurate measurements such as the ones used here the 'thermal' tails to the lines are detectable and should strictly be excluded from any analysis of extinction. Hall and Williamson deliberately excluded these tails and judging by the agreement between calculated and observed intensities from cold worked α-brass (extinction free), it appears that Averbach and Warren excluded them also.

In attempting to compare the integrated intensities of lines from cold worked and annealed metals respectively a similar difficulty arises. The long tails due to cold work certainly overlap for adjacent lines. The slight rise in background level observed in Hall and Williamson's figure 1 and in Averbach and Warren's figure 3, makes it reasonable to suppose that there is extensive overlap over the entire spectrum. Ideally, this background should be resolved into the component tails and these tails included with the line intensities before comparison. Such a procedure was in fact attempted by Averbach and Warren, but it is difficult to justify their (or any other) method of resolving tails. Hall and Williamson on the other hand used total increase in background intensity and total line intensity so that this problem did not arise, and found that within experimental accuracy there was no change in total reflected intensity. The rise in background level is itself very small, being very little greater than the statistical scatter of the experiments so that many more measurements of line intensity and background level are really necessary before results given here can be regarded with certainty.

Confirmation of the theory that cold work produces no change in total reflected intensity has been given by Weiss (1952 b) from a neutron diffraction study of cold work in α-brass. In this case extinction is said to be much smaller than in the present case and Weiss showed that there was no significant change in the total reflected intensity. He also showed that there was no change in the total integrated intensities of the two adjacent lines 3 and 4 on cold work. These intensities, however, were measured relative to the background level of the annealed sample, a level significantly lower than that near the two reflections from the cold worked state. These results in no way conflict with those given in this paper because of the differences in the definition of integrated intensity (discussed above). There is now universal agreement that, after correction for extinction, there is no change in total reflected intensity with cold work and hence the slight fall in integrated intensity reported in this paper corresponds to a slight increase in the intensity of the tails remote from the lines, an observation in good agreement with dislocation theory.

ACKNOWLEDGMENTS

We are grateful to Drs. Weiss and Lang for their helpful criticisms and to Dr. Hirsch for valuable discussions.

We are indebted to the University of Birmingham for generous financial support and to the late head of the Department of Metallurgy, Professor D. Hanson, for his encouragement. Dr. H. M. Finniston gave valuable encouragement during the later stages of this work, which was carried out at the Atomic Energy Research Establishment, Harwell.

REFERENCES

AVERBACH, B. L., and WARREN, B. E., 1949, *J. Appl. Phys.*, **20**, 1066.
BRINDLEY, G. W., 1940, *Proc. Phys. Soc.*, **52**, 117.
GAY, P., HIRSCH, P. B., and KELLY, A., 1953, *Acta Metallurgia*, **1**, 315.
GUINIER, A., 1951, *X-Ray Crystallographic Technology* (London : Hilger and Watts).
HALL, W. H., and WILLIAMSON, G. K., 1951 a, *Proc. Phys. Soc.* B, **64**, 937; 1951 b, *Ibid.*, **64**, 946.
LANG, A. R., 1953, *Proc. Phys. Soc.* B, **66**, 1003.
SMALLMAN, R. E., 1953, *Thesis* (University of Birmingham).
WEISS, R. J., 1952 a, *Proc. Phys. Soc.* B, **65**, 553 ; 1952 b, *J. Appl. Phys.*, **23**, 1379.

The Crystal Structure of Lanthanum–Strontium Manganites

By M. G. HARWOOD
Philips Electrical Limited, Material Research Laboratory,
New Road, Mitcham Junction, Surrey

MS. received 25*th January* 1955 *and in final form* 18*th February* 1955

Abstract. The lanthanum manganite–strontium manganite system has been examined as a function of the lanthanum content. The perovskite phase is stable from 23% to 100% lanthanum. At the lower limit of stability it is of cubic symmetry the unit cell becoming rhombohedral between 50% and 60% lanthanum.

In the lanthanum manganese oxygen system of general formula $LaMnO_x$ the value of x has been varied between 3·24 and 3·07. Over the entire range the structure is of the perovskite type with small type of cell. At the oxygen rich end of the system the symmetry is rhombohedral while at the other end it is cubic.

In both systems the relations between the lattice parameters and composition and magnetic Curie temperatures are indicated.

§ 1. INTRODUCTION

THE perovskite structure has provided much interest for many years now from both academic and dielectric points of view. More recently magnetic (Jonker and van Santen 1950) and luminescent (Klasens, Zalm and Huysman 1953) materials having this structure have been studied. The field of applications is likely to be still further widened so that a better knowledge of this apparently simple structure is desirable in the interests of material development.

The work reported here is concerned with the manganites of lanthanum with and without the addition of strontium. Departure from stoichiometry has been given consideration and its effect on the structure reported. In the $LaMnO_3$–$SrMnO_3$ system, Jonker and van Santen have determined the structure in the region of 40–50% $SrMnO_3$. In this range they state that these materials are cubic but outside it they are not so.

Other structural work bearing on this subject has been done by Askham, Fankuchen and Ward (1950) who have found that $LaCoO_3$ has a perovskite structure and a rhombohedral unit cell. They made no attempt to vary the oxygen content.

§ 2. SAMPLES

The compositions studied are divided into two groups. One involves the system (La_ySr_{1-y}) MnO_3 where $1 \geqslant y > 0{\cdot}2$ and the other $LaMnO_x$ where $3{\cdot}25 > x > 3{\cdot}05$.

The preparation of the samples involved normal ceramic techniques. This included the mixing of the oxides or carbonates, pressing into pellets, prefiring at about 1000°c, crushing, remixing and refiring. The final firing lay between 1250°c and 1650°c in an atmosphere of oxygen, air or nitrogen according to the amount of oxygen finally required in the sample.

§ 3. EXPERIMENTAL TECHNIQUE

3.1. *X-ray Diffraction*

The pellets were crushed and rolled into rod samples about ½ mm diameter with the aid of a suitable binder. Most photographs were taken in Debye–Scherrer type cameras 11·5 cm in diameter with iron radiation and a manganese filter, and a few with chromium radiation and a vanadium filter when higher dispersion was required. Almost all exposures were made at room temperature but a few were taken when the specimen was cooled with liquid oxygen or heated to about 400°c.

3.2. *Magnetic Measurements*

As only a rough measure of the Curie temperature was required a simple arrangement was constructed. This consisted of a piece of 'Ticonal' in the form of a hollow cylinder about half an inch long and rather less in diameter. Into this was cemented a thermocouple whose tip was located in the end plane of the cylinder. A sample, magnetic at room temperature, was fixed to this end and slowly heated in a furnace. When the Curie temperature was reached it fell off and dropped on to an alarm system. The temperature as indicated by the thermocouple was recorded.

For samples not magnetic at room temperature a small vessel was filled with liquid oxygen. Into this the sample was dropped and soon after the magnet was also lowered into it. The sample was picked up by the magnet and was allowed to warm up slowly in air free from draughts. Again when the Curie point was reached the alarm indicated that the temperature had to be recorded. The measurements were carried out on different pieces of material at each composition and a standard deviation of the temperature readings was found to be about ± 2°c which was quite adequate for the purpose.

§ 4. INTERPRETATION OF DIFFRACTION PATTERNS

All the compounds examined were of the perovskite type and all but two of these were either of cubic or rhombohedral symmetry. The two exceptions were orthorhombic.

Figure 1 (Plate) shows the patterns obtained from the two different types of structure. The upper one, being obviously cubic, calls for no comment. The lower was found to be rhombohedral. This was determined in the following manner. A split in the (111) reflection cannot be obtained from an orthogonal unit cell: this eliminates cubic tetragonal and orthorhombic symmetry. For a monoclinic unit cell and the sides nearly equal, a symmetrical doublet should occur: this is seen not to be the case. A rhombohedral cell demands a doublet consisting of $(11\bar{1}+1\bar{1}1+\bar{1}11)$ as one member and (111) as the other. If $\alpha < 90°$ the stronger line will appear at the larger Bragg angle while conversely if $\alpha > 90°$ the arrangement found in the lower picture will be obtained.

The (210) reflections consist of a symmetrical doublet. Tetragonal symmetry requires three lines of equal intensity while a monoclinic unit cell would give a symmetrical triplet with the middle line strongest. Rhombohedral symmetry on the other hand requires a doublet with lines of equal intensity as is found in the photograph.

In this way all reflections were indexed satisfactorily. The lattice parameters were then computed on the basis of successive approximations and the results plotted against composition as a variable.

§ 5. Results

5.1. *$LaMnO_3$–$SrMnO_3$ System*

A summary of the crystallographic and magnetic measurements is given in table 1 and also graphically in figures 2 and 3.

Table 1

La/Sr	Structure	Unit cell parameters		Curie temp.
		a (Å)	α	(°c)
94 : 6	Rhombohedral	3·883	90° 35·4′	−62
86 : 14	Rhombohedral	3·881	90° 29·4′	+16
72 : 28	Rhombohedral	3·8725	90° 25·8′	+74
68 : 32	Rhombohedral	3·870	90° 24′	+85
65 : 35	Rhombohedral	3·866$_5$	90° 19·2′	+94
60 : 40	Rhombohedral	3·861	90° 18·6′	+81
52 : 48	Cubic	3·858	90° —	+71
34 : 66	Cubic	3·832	90° —	+68
22 : 78	Cubic	3·824	90° —	+61

Figure 2 shows the variation of the structure with the lanthanum–strontium ratio. All samples in this series were made under the same firing schedule and the results relate to this rather than to the constancy of oxygen content. The latter was subsequently found to have varied somewhat through the series but was not the subject of detailed study. At low strontium content the unit cell is rhombohedral, both parameters decreasing with increasing strontium. Between 40 and 48% strontium a transition to cubic symmetry occurs, the lattice parameter continuing to decrease after this. The transition from one form to the other has not been explored in detail but it appears to be somewhat abrupt.

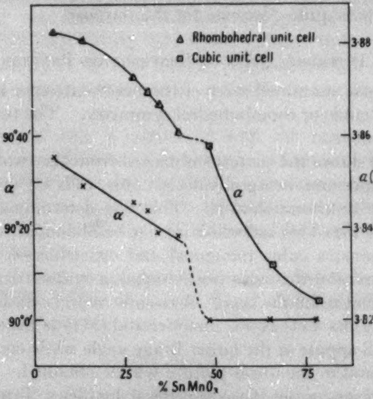


Figure 2.

The Curie temperature was measured on the same samples and the results are seen in figure 3. A maximum for one structure will be seen. The shape of the curve is similar to that obtained by Jonker and van Santen. It appears from the change in direction at the structural transition that the latter has an effect on the Curie temperature.

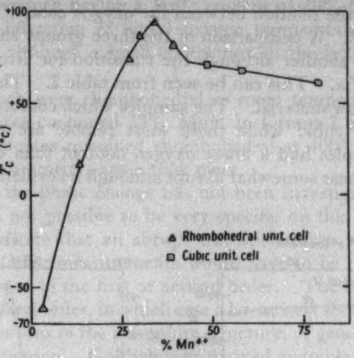


Figure 3.

5.2. *LaMnO$_3$ System*

The samples in this sytem were divided into three groups as three batches of MnO_2 were used. The parameter x was varied between 3·05 and 3·25. Attempts to make stoichiometric $LaMnO_3$ proved fruitless, but the use of a reducing atmosphere was not tried.

Table 2 summarizes all results in this system, some of which are also represented graphically in figures 4 and 5.

Table 2

	Structure	Unit cell parameters		Curie temp. (c°)	Firing conditions	Batch of MnO_2
x		a (Å)	α			
3·07	Cubic $^{1/3}$	3·904	90° —	−97	1400°c in N_2 (2 hours)	1
	Cubic $^{2/3}$	3·885	90° —	−97		
3·10	Rhombohedral	3·880	90° 32·2′	−110	1400°c in air (2 hours)	
3·17	Rhombohedral	$3{\cdot}83_5$	90° 31·2′	+6	1000°c in air (2 hours)	
3·13	Cubic	3·961	90° —	−115	1650°c in air (4 hours)	2
3·15	Cubic	3·889	90° —	−112	1400°c in N_2 (4 hours)	
3·17	Rhombohedral	3·885	90° 34·8′	−55	1250°c in N_2 (4 hours)	
3·22	Rhombohedral	3·879	90° 33·0′	−112	1400°c in O_2 (4 hours)	
3·23	Rhombohedral	3·878	90° 31·8′	−108	1250°c in O_2 (3½ hours)	
3·146	Cubic	3·896	90° —	−95	1250°c in N_2 (4 hours)	3
3·152	Rhombohedral	3·879	90° 35·4′	−67	1400°c in O_2 (4 hours)	
3·156	Rhombohedral	3·878	90° 46·8′	−63	1250°c in O_2 (4 hours)	

Figure 4 gives the relation between the oxygen content and the lattice parameters of Group 2. A comparison of the three groups shows that similarities exist between one another although the transition for Group 1 is displaced to the lower values of x. This can be seen from table 2. This may be due to the impurities in the raw material. The samples which conform most nearly to the ideal formula are cubic while those most remote are rhombohedral. The orthorhombic samples had a lower oxygen content than even the cubic ones. The transitions appear somewhat abrupt although no evidence for a discontinuity can really be seen.

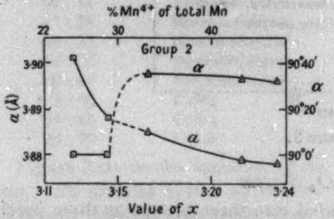


Figure 4.

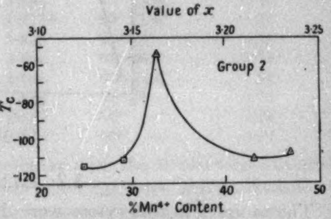


Figure 5.

No co-existence of the two forms could be observed although special attention was paid to this possibility.

The relation between the Curie temperature and the Mn^{4+} concentration is given in figure 5 for batch 2. Although the number of points in the curve are rather few the character of the relation for each group is similar.

In spite of the apparent displacement of the (lattice parameter, oxygen content) relation of Group 1 relative to Groups 2 and 3 the relations between Mn^{4+} concentration and Curie temperature brings the three groups into line with one another, although the structural transition in Group 1 still appears at a different Mn^{4+} concentration from that in the other two groups.

§ 6. DISCUSSION

The transition from a cubic directly to a rhombohedral modification of the perovskite structure has been reported only once before (Shirane and Hoshino 1952). That is in the system $(Pb . Ba)ZrO_3$. There the cubic phase is paraelectric and the change to rhombohedral, effected either by reducing the temperature or by increasing the lead content, is associated with a ferroelectric Curie point. Here in common with other ferromagnetic materials the phase change is not connected with the appearance or disappearance of the spontaneous magnetization. This will be clearly seen in figures 3 and 5. In figure 3 the phase change takes place at about 45% Mn^{4+} at room temperature whereas the Curie temperature is about 70°c. In figure 5 all samples will be seen to be non-ferromagnetic at 20°c, some being cubic and others rhombohedral.

The temperature sensitivity of the structure was further examined on three samples which have compositions very near that involving the phase change. These, it was considered, should be more sensitive to heating or cooling than those lying more remote in a compositional sense from the transition point.

In particular a specimen having a Mn^{4+} content of 29% from Group 2 was examined by x-ray diffraction at liquid oxygen temperature. No change in symmetry was noted although a small contraction in the lattice parameter was found.

Two samples which were rhombohedral at room temperature were heated to about 400°c. These contained 33% Mn^{4+} in Group 2 and 31% Mn^{4+} in Group 3. Both specimens remained rhombohedral at 400°c although a small change in the lattice parameters again took place.

The character of the phase change has not been investigated and from the results obtained it is not possible to be very specific on this point. However, figure 2 seems to indicate that an abrupt transition takes place, as also does figure 4, Group 2. Other measurements would have to be made to determine whether the transition is of the first or second order. The indications are that it might be one of the first order, in which case a latent heat should be measurable.

The packing of the ions in the perovskite structure, of general formula ABO_3, has received much attention. Goldschmidt derived a simple geometric relation which involves the A ion being in contact with the oxygen ions as nearest neighbours and also the B ion being in contact with the six anions surrounding it. This leads to the relation $R_A + R_O = \sqrt{2}\,(R_B + R_O)$ where R_A, R_O and R_B are the radii of the A, oxygen, and B ions respectively. As the lattice will tolerate certain departures from these values Goldschmidt introduced a tolerance factor t on the right-hand side of the above equation. This factor has been calculated for the La–$SrMnO_3$ system. For this purpose the ionic radii used were those derived empirically by Goldschmidt and corrected for the effect of coordination relative to oxygen as computed by Pauling. They are:

Ion	Coordination	Radius (Å)
Sr^{2+}	12	1·37
La^{3+}	12	1·30
Mn^{3+}	6	0·70
Mn^{4+}	6	0·52
O^{2-}	6	1·32

The results are indicated in figure 6. The compounds with a tolerance factor near to unity are cubic, while those more remote are rhombohedral. The transition occurs at a tolerance factor of 0·96–0·97. It is interesting that the structure is stable up to a tolerance factor of 1·008 and that the lattice at this value is cubic. Barium titanate, another compound with a tolerance factor above unity, is, of course, tetragonal at room temperature.

The relation between the Curie temperature and the tetravalent manganese content has a maximum for each material studied at about 30–35% Mn^{4+}. Experimental evidence is not sufficient to show whether this is due to the negative interactions between the Mn^{4+}–Mn^{4+} balancing the positive Mn^{3+}–Mn^{4+} interactions or whether an ordering mechanism is set up having a maximum degree of order at that Mn^{4+} content. The possibility of antiferromagnetic domains thus being formed is a real one and neutron diffraction would reveal this.

§ 7. CONCLUSIONS

The lanthanum–strontium manganite system has a perovskite structure when the lanthanum content is in the range 23–100%. It is cubic at the lower and rhombohedral at the higher limits of stability. The lanthanum manganese oxygen system has a perovskite structure for the whole range of oxygen contents which were examined. The samples conforming more closely to the ideal ABO_3 formula are cubic while the others are rhombohedral.

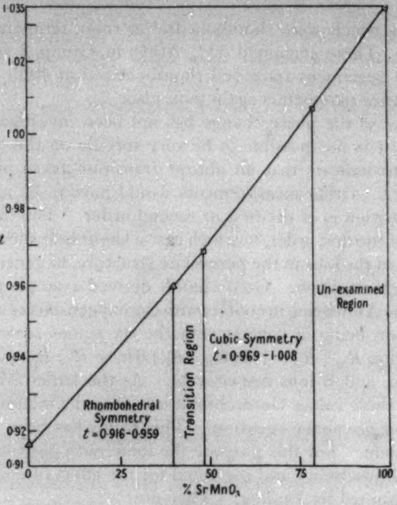


Figure 6.

The relationship between Curie temperature and tetravalent manganese content gives a maximum value for each system at 30–35% Mn^{4+}. This must be explained in any theory which attempts to account for the magnetic interactions of the ions in these systems. The change of structure from cubic to rhombohedral is somewhat abrupt, but evidence is not sufficiently complete to show whether it is discontinuous or not. The (Curie temperature, Mn^{4+}) curve in the La–$SrMnO_3$ system shows that the magnetic interactions are structure sensitive.

ACKNOWLEDGMENTS

I wish to thank Mr. J. A. M. van Moll and the Directors of Philips Electrical Limited for permission to publish this paper.

REFERENCES

Jonker, G. H., and van Santen, J. H., 1950, *Physica*, **16**, 337.
Klasens, H. A., Zalm, P., and Huysman, F. O., 1953, *Philips Res. Rep.*, **8**, 441.
Askham, F., Fankuchen, I., and Ward, R., 1950, *J. Amer. Chem. Soc.*, **72**, 3799.
Shirane, G., and Hoshino, S., 1952, *Phys. Rev.*, **86**, 248.

The Infra-red Properties of Some Metallic Films

By J. N. HODGSON†
Department of Physics, The University, Reading, Berks.

Communicated by D. E. Blackwell; MS. received 24th February 1955 *and in amended form* 18*th April* 1955

Abstract. The values of the refractive index *n* and the extinction coefficient *k* have been measured for evaporated films of gold, silver, copper, zinc, tin and aluminium, at wavelengths between 1 and 15 μ, by a reflection method. The apparatus for the measurements contained two selenium film polarizers, and an infra-red spectrometer with two rock salt prisms and a radiation thermocouple. Theoretical formulae for the infra-red properties of a metal are discussed and it is shown that the present experimental results are not sufficiently accurate to distinguish between the classical free electron theory and later modifications. The experimental results have been compared with theory by a graphical method. Some inferences about electronic band structures are made from the comparison of experimental points with theoretical curves.

§ 1. Introduction

The optical properties of metals have been widely investigated since Maxwell's formulation of the theory of electromagnetic waves in absorbing media. Using Maxwell's theory, *n* the refractive index and *k* the extinction coefficient can be found by measurements of the change in polarization of light reflected from the surface of a metal. The original theory shows that in the limit as the frequency $f \to 0$, then $nkf \to \sigma_0$ and $n^2 - k^2 \to \epsilon_0$, where σ_0 and ϵ_0 are respectively the conductivity and dielectric constant for static fields. This theory gives no more information about the variation of *n* and *k* with frequency. To make predictions about this variation, one must turn to theories of the atomic structure of metals.

Drude (1904) has deduced expressions for *n* and *k* as a function of frequency from the classical free electron model of a metal. This theory gives the values of *n* and *k* in terms of the frequency and two parameters, the density of free electrons and their relaxation time. Kronig (1931) has shown that Drude's theory is valid for a free electron model obeying quantum mechanics, with a re-interpretation of the parameters. The free electron formulae for *n* and *k* should be valid for frequencies below the threshold for transitions between electronic bands. This threshold frequency occurs for most metals in the visible or near infra-red. The optical properties of most metals in the visible spectrum are therefore influenced by inter-band transitions. The infra-red properties of metals, at least for wavelengths greater than 5 μ, may be expected to obey the formulae of the free electron theory. These formulae have been modified by Dingle (1953) to allow for the anomalous skin effect.

---

† Now at Department of Physics, University College of North Staffordshire, Keele, Staffordshire.

This paper describes measurements of *n* and *k* for some metallic films at wavelengths between 1 and 13 μ. It seems that the comparison of the experimental results with the free electron theory can give useful information about the electronic band structure of a metal.

§ 2. MEASUREMENT OF *n* AND *k* IN THE INFRA-RED

The most accurate methods of measuring *n* and *k* in the visible spectrum have made use of phase compensators. These compensators are usually restricted by absorption to wavelengths less than 2·5 μ. Only simple polarizers are available for polarization measurements between 2·5 and 15 μ. There are several reflection methods for measuring *n* and *k* which require only one or two polarizers. Ingersoll (1910) measured *n* and *k* for some bulk metals at wavelengths between 0·65 μ and 3 μ. He calculated *n* and *k* from the measured values of the principal angles of incidence $\bar{\theta}$ and azimuth $\bar{\psi}$. His method has the advantage that it uses a radiation receiver only to indicate equality of two intensities. Similar methods of finding *n* and *k* in the infra-red by measurements of $\bar{\theta}$ and $\bar{\psi}$ have been used by Forsterling and Freedericksz (1913) and by Weiss (1948). Values for *n* and *k* can be found by measurements of reflectivity for at least two values of θ, the angle of incidence. This method has been used by Bueche (1948) for some metallic films at wavelengths between 1 μ and 2·5 μ. The method is not capable of high accuracy because the variation of reflectivity with θ is not sensitive to changes in *n* and *k*.

The principal angles method is not suitable for measurements on most metals at wavelengths greater than 2·5 μ, because of the large values of $\bar{\theta}$ which occur. The value of $\bar{\theta}$ for silver, for example, increases steadily with wavelength in the infra-red, reaching 88° at about 3 μ. The use of large angles of incidence involves large reflecting films to counteract foreshortening, and small angular dispersion of the incident beam to give a well defined value of θ. The maximum value of θ with the apparatus used for the present work was 88°, which would limit measurements by the principal angles method on silver, for example, to wavelengths below 3 μ.

The method used for the present measurements depended on the analysis of elliptically polarized radiation reflected from a metallic film. The incident radiation was plane polarized at 45° to the plane of incidence. It was reflected by a metallic film at an angle of incidence θ, into a second polarizer. As the second polarizer was rotated, the transmitted intensity varied sinusoidally. The phase and amplitude of this variation gave the elements of the vibration ellipse, from which *n* and *k* can be calculated. This method has the advantage that it requires measurements at only one, arbitrary, value of θ. The most accurate results are obtained when θ is near to $\bar{\theta}$. This means that for most metals, when $\lambda > 3\,\mu$, θ should have the maximum value of 88°. Since the shape of the vibration ellipse is derived from an intensity variation, a receiver of known intensity response is required.

The reflected vibration ellipse may be specified by its axes *a* and *b* ($a \geqslant b$), and the acute angle γ between the major axis and the normal to the plane of incidence. It is convenient to introduce an angle η, defined by:

$$\tan\eta = b/a \qquad (0 \leqslant \eta \leqslant 45°). \qquad \ldots\ldots(1)$$

In the calculation of n and k, the values of ψ the azimuth and Δ the phase angle are first calculated from η and γ. The necessary formulae can be derived by geometrical consideration of the vibration ellipse:

$$\cos 2\psi = \cos 2\eta \cos 2\gamma ; \qquad \tan \Delta = -\tan 2\eta \operatorname{cosec} 2\gamma. \qquad \ldots\ldots(2)$$

Formulae for calculating n and k from ψ and Δ have been given by Price (1946):

$$k^2 - n^2 = \frac{2\sin^2\theta \tan^2\theta(\cos\Delta + \sin 2\psi)}{\sin 2\psi\,(\cos\Delta + \operatorname{cosec} 2\psi)^2} - \tan^2\theta \qquad \ldots\ldots(3)$$

$$2nk = \frac{2\sin^2\theta \tan^2\theta \sin\Delta}{\tan 2\psi(\cos\Delta + \operatorname{cosec} 2\psi)^2}. \qquad \ldots\ldots(4)$$

The intensity $I(\phi)$ transmitted by the second polarizer as a function of ϕ, the acute angle between the transmitted vibration of the polarizer and the normal to the plane of incidence, is given by

$$CI(\phi) = (a^2 + b^2) + (a^2 - b^2)\cos 2(\phi - \gamma) \qquad \ldots\ldots(5)$$

where C is a constant. The corresponding output $D(\phi)$ from the radiation receiver, is given by

$$D(\phi) = T(\phi)I(\phi) \qquad \ldots\ldots(6)$$

where $T(\phi)$ allows for the instrumental polarization. The values of $T(\phi)$ were found by measurements in which the radiation was allowed to pass directly between the polarizers, without reflection at a metallic film. To find a, b and γ by equation (5), values of $I(\phi)$ for at least three values of ϕ are required. Experimental measurements of $I(\phi)$ at a larger number of values of ϕ showed that the intensity variation agreed with (5) within the experimental error. In most of the measurements $D(\phi)$ was measured for three values of ϕ at intervals of 45°.

§ 3. THE APPARATUS

An apparatus for reflection measurements in the infra-red between $1\,\mu$ and $15\,\mu$, was constructed with two selenium film polarizers and a rock salt prism spectrometer. The infra-red properties of metals do not show any sharp changes with wavelength, so a low spectral resolution in the spectrometer was sufficient. It was an advantage to use as small a resolution as adequate, so as to have as much intensity as possible at the thermocouple. Because of the large changes in n and k between $1\,\mu$ and $15\,\mu$ an appreciable intensity of scattered, short wavelength radiation could falsify the results at longer wavelengths. It was therefore necessary to ensure that a negligible intensity of scattered radiation reached the thermocouple.

A spectrometer with two 60° rock salt prisms and a Hilger vacuum thermocouple was used. The spectrometer prisms were arranged so that their dispersions cancelled. The resolving power of the spectrometer was controlled chiefly by the intermediate slit width. The results of previous workers indicated that a resolving power of about 20 would be adequate for the present work. The intermediate slit width was adjusted at each wavelength to give a resolving power of 20 to 30. The intensity of scattered radiation of λ less than $2{\cdot}5\,\mu$ was found to be less than 0·5%of the total intensity at $\lambda = 15\,\mu$.

The collimating system was designed for angles of incidence up to 88° with 3 in. × 1 in. metallic films, deposited on microscope slides. The arrangement is shown in figure 1. The source of radiation was a Nernst glower, connected to the mains through a barretter. The radiation from the glower was focused by a mirror on to a slit 1·5 mm × 7 mm followed by the first polarizer. A second mirror M_2 formed a three-fold magnified image of the slit at the centre of the metallic film. This film was mounted on a turntable with its short edge vertical. A third mirror focused the reflected beam on to the entrance slit of the spectrometer, preceded by the second polarizer. The angles of incidence at the collimating mirrors were too small to introduce appreciable polarization. The range of values of θ among the rays incident on a film, was equal to the angle subtended at the film by the mirror M_2. Not all the rays from M_2 reached the thermocouple however, because of the limited aperture of the spectrometer prisms. The effective rays were located by moving a screen with a vertical edge across M_2 and noting the response at each position of the screen. This experiment showed that the range of θ was 0·6°.

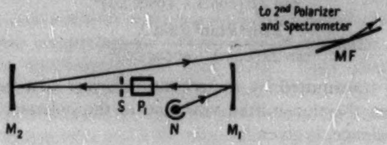


Figure 1. N, Nernst glower ; P_1, first polarizer ; S, slit ; MF, metallic film ; M_1, M_2, aluminized concave mirrors.

The thermocouple was connected to a mirror galvanometer. The deflection of this galvanometer was measured by a photocell amplifier with negative feedback. Measurements showed that the amplifier plus thermocouple had an intensity response linear to ± 0·5%.

Selenium film polarizers were used, similar to those described by Elliot *et al.* (1948). Six selenium films were mounted on a piece of one inch aperture brass tube, at an angle of 22° to the axis of the tube. Two of these polarizers were mounted in converging beams of radiation, near the spectrometer entrance slit and the slit S (figure 1). The semi-angle of the beams was 3°. The efficiency of the polarizers was estimated from the minimum and maximum intensities when radiation was passed through both polarizers, one of which was rotated. The transmission for the allowed vibration was 0·69 and the degree of polarization of the transmitted radiation was 0·994. The low transmission may be due to the rather wrinkled shape of some films in the polarizers. The small unpolarized component transmitted by the polarizers necessitated a correction to the observed amplitude ratio to give the ellipticity b/a.

The metallic films used for measurements were made by vacuum evaporation in an apparatus of standard design. An electrical discharge in the bell jar was used to give a final cleaning to the slides on which the metallic films were deposited.

§ 4. EXPERIMENTAL ERRORS

The values of n and k were affected by errors in the measurement of intensities, in the value of θ and in the setting of the second polarizer for $\phi = 0$ (see equation (5)). A constant error in either of the last two quantities would cause systematic errors in n and k. To detect any such errors, measurements were made at each wavelength for two values of θ. By trial variations in the assumed values of θ and the zero of ϕ, it was found that the small variations with θ of the apparent values of n and k were not due to constant errors in θ or ϕ.

The use of a straight mean for the value of θ in equations (3) and (4) will lead to certain errors, especially for large values of θ. The measured values of γ and η are certain averages over the bundle of rays reflected at the film. Owing to the complexity of the equations, it has not been possible to find a formula for the effective value of θ for a bundle of rays. An estimate of the possible error due to taking a straight mean for θ has been made from (3) and (4). For values of θ near 90°, k^2-n^2 and nk are nearly proportional to $\tan^2\theta$. If the difference between the mean value and the effective value of θ is $\delta\theta$, then the error in k^2-n^2 is $\delta(\tan^2\theta)$. The semi-angle of the bundle of incident rays was 0·3°. Taking $\delta\theta=0{\cdot}1°$, we find for $\theta=84°$, 86°, 88°, $\delta(\tan^2\theta)/\tan^2\theta=3\%$, 5%, 10% respectively. A comparison of measurements for $\theta=86°$ and 88°, and the agreement of the experimental points for k^2-n^2 for a silver film, with a theoretical curve, both indicate that the error was less than 10% for $\theta=88°$. The reason for the greater scatter of points for nk of the silver film is discussed in the next paragraph.

The probable error in γ and η due to random errors in measurements of intensity was estimated at ±0·5°. The effect of these errors on the values of n and k depends on the values of γ and η. For films of silver, gold and copper some simplifying approximations may be made. For these metals in the infra-red, 2ψ is always between 80° and 90°. If we put $2\psi=90°-2\chi$, equations (3) and (4) can be simplified for the case when χ is small. To the first order, k^2-n^2 is independent of χ, and nk is proportional to χ. The probable error in χ is about the same as in γ or η, i.e. ±0·5°. The errors in the values of nk for these metals are therefore large, being ±10% and ±20% when 2ψ is 80° and 85°, respectively. This probably accounts for the greater scatter of the points for $nk\nu$ of a silver film, about the theoretical curve, compared with the points for k^2-n^2.

§ 5. FREE ELECTRON THEORY

The theory of Drude (1904) for the infra-red properties of a metal will be given in a notation similar to that of Mott and Zener (1934). Drude found expressions for n and k in terms of the density of free electrons N_0 and their relaxation time τ. The formulae take a simple form when one introduces

$$\nu_R=1/2\pi c\tau;\qquad \nu_0^2=N_0e^2/m\pi c^2. \qquad \ldots\ldots(7)$$

ν_0 and ν_R have the dimension of wave numbers and will be given in cm^{-1} in the rest of this paper. The formulae for n and k at wave number ν are

$$k^2-n^2+1=\nu_0^2/(\nu^2+\nu_R^2) \qquad \ldots\ldots(8)$$

$$2nk\nu=\nu_0^2\nu_R/(\nu^2+\nu_R^2). \qquad \ldots\ldots(9)$$

The static conductivity σ_0 is given in e.s.u. by

$$\sigma_0=\tfrac{1}{2}c\nu_0^2/\nu_R. \qquad \ldots\ldots(10)$$

Kronig (1931) has used quantum mechanics to find expressions for n and k for a metal at low frequencies. His formulae show the same frequency dependence as (8) and (9) but the parameters have slightly different interpretations.

The theory of the anomalous skin effect, as originally given by Reuter and Sondheimer (1948), indicated that (8) and (9) would be valid for the frequencies and temperatures of the present measurements. A parameter p was introduced by these authors, to denote the fraction of electrons specularly reflected on striking the surface of a metal. Their calculations for high frequencies were limited to the case $p=1$. Holstein (1952) and Dingle (1952) later showed that when $p<1$ a constant deviation from the classical formulae will persist at all frequencies and temperatures. Experimental results of Chambers (1952) on the anomalous skin effect at radio frequencies indicate that usually $p \simeq 0$. Dingle (1953) has published calculations of the anomalous skin effect for any value of p and wide ranges of frequency and temperature. His formulae take the form of series expansions valid for certain ranges of ν. Dingle's formulae (4·5) and (4·6), with $p=0$, (p. 734), are given below in the present notation:

$$k^2-n^2=\nu_0{}^2\nu^{-2}-\nu_0{}^2\nu_{\mathrm{R}}{}^2(1+1{\cdot}08x+0{\cdot}31x^2)\nu^{-4}+\dots \qquad \dots\dots(11)$$

$$2nk\nu=\nu_0{}^2\nu_{\mathrm{R}}(1+0{\cdot}433x)\nu^{-2}-\nu_0{}^2\nu_{\mathrm{R}}{}^3(1+1{\cdot}89x+1{\cdot}25x^2+0{\cdot}28x^3)\nu^{-4}+\dots \qquad \dots\dots(12)$$

$x=$ Dingle's $\sigma=\sqrt{3}u\nu_0/2\mathbf{c}\nu_{\mathrm{R}}$, where u is the electronic velocity at the surface of the Fermi sphere; ($u=[3\nu_0{}^2\boldsymbol{h}^3\mathbf{c}^2/8\boldsymbol{m}^2\mathbf{e}^2]^{1/3}$). The first two terms of each series are good approximations to the sums if $\nu \gg \nu_{\mathrm{R}}$ and $x \leqslant 1$. When (8) and (9) are expanded in powers of $\nu_{\mathrm{R}}\nu^{-1}$, the first two terms in each series are just those outside the brackets in (11) and (12), (neglecting 1 in (8)). The accuracy of the experimental values for n and k was not sufficient to distinguish between the variations with ν implied by (8), (9) and (11), (12), in the region $\nu \gg \nu_{\mathrm{R}}$. The experimental data had been compared with (8) and (9) before Dingle's calculations appeared. The only useful correction to this comparison, with the present experimental error, was in the value of ν_{R}. We see from (12) that ν_{R} in (9) should be replaced by

$$\nu_{\mathrm{R}}{}'=\nu_{\mathrm{R}}(1+0{\cdot}433x)=\nu_{\mathrm{R}}+0{\cdot}375u\nu_0/\mathbf{c}. \qquad \dots\dots(13)$$

The second term in (11) is almost equal to $\nu_0{}^2\nu_{\mathrm{R}}{}'^2\nu^{-4}$, so the two pairs of formulae predict nearly the same variation with ν, when the change in ν_{R} is allowed for.

§ 6. COMPARISON OF EXPERIMENT AND THEORY

The experimental values of k^2-n^2 and nk for each metallic film have been compared with expressions of the form of (8) and (9), with ν_{R} replaced by $\nu_{\mathrm{R}}{}'$ to allow for the anomalous skin effect. The parameters ν_0 and $\nu_{\mathrm{R}}{}'$ were chosen for the best fit with experimental values at long wavelengths. For some of the metals investigated the influence of inter-band transitions extended into the infra-red. Formulae (8) and (9) are not valid for these metals in the near infra-red. The best values of ν_0 and $\nu_{\mathrm{R}}{}'$ for each film were found by a graphical method. Formulae (8) and (9) are equivalent to

$$X=(1+k^2-n^2)^{-1}=\nu_0{}^{-2}(\nu^2+\nu_{\mathrm{R}}{}'^2) \qquad \dots\dots(14)$$

$$Y=(2nk\nu)^{-1} \quad =\nu_0{}^{-2}\nu_{\mathrm{R}}{}'^{-1}(\nu^2+\nu_{\mathrm{R}}{}'^2). \qquad \dots\dots(15)$$

The experimental points for X and Y were plotted on a graph against ν^2. According to (14) and (15), these points lie on two straight lines, intersecting on the ν^2 axis at $\nu^2 = -\nu_R'^2$ and with slopes in the ratio $\nu_R' : 1$. The graph for a gold film is shown in figure 2. Because of the small value of ν_R', the intersection on the ν^2 axis does not give an accurate value of ν_R' for gold.

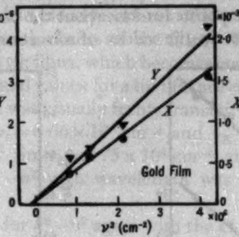


Figure 2. $X=(k^2-n^2+1)^{-1}$; $Y=(2nk\nu)^{-1}$.

The experimental values of the parameters for several metallic films are given in the table. ν_0 was calculated from the slope of the graph of X plotted against ν^2, according to (14); n_0 the number of free electrons per atom was calculated from ν_0 by (7); n_0(a.s.e.) are values found by Chambers (1952) from the anomalous skin effect at radio frequencies; ν_R was derived from ν'_R and ν_0 by (13), and the static conductivity σ_0 from (10).

Metal	ν_0(cm^{-1})	ν_R(cm^{-1})	n_0 (i.r.)	n_0 (a.s.e.)	σ_0 (i.r.) (e.s.u.)	σ_0 (b.) (e.s.u.)	$0{\cdot}375\ \nu_0 u/c$ (cm^{-1})
	$\times 10^4$	$\times 10^3$			$\times 10^{16}$	$\times 10^{16}$	$\times 10^3$
Au	5·21	0·30	0·53	0·6	13·6	41·3	0·072
Ag	7·55	0·37	1·08		23·1	61·4	0·135
Cu	6·81	0·72	0·61		9·7	51·8	0·113
Zn	3·43	0·53	0·21		3·3	15·8	0·036
Sn	6·66	1·31	1·33	1·1	5·1	8·28	0·109
Al	8·64	1·53	1·35	2·2	7·3	31·6	0·169

σ_0 (b.) = bulk at 18°c.

§ 7. Experimental Results

(i) *Gold.* The method of evaporating gold from a molybdenum boat, as used by Weiss (1948), was found unsatisfactory. Molten gold reacted with the molybdenum and films produced had widely variable properties. An evaporating source of a crucible of alundum cement on a tungsten wire base was found to produce gold films with reproducible properties. The experimental values for a gold film are compared with theoretical curves in figure 3. The points at long wavelengths agree with the curves within the experimental error. The experimental values for nk deviate from the curve for λ less than $2{\cdot}5\,\mu$, as is clearly shown by the more detailed results of Weiss and Forsterling. This deviation is presumably due to inter-band transitions.

(ii) *Silver*. Evaporation from a molybdenum boat was found to give satisfactory films of silver. The properties of a silver film were measured immediately after deposition to minimize the effect of atmospheric corrosion. The experimental values are compared with theoretical curves in figure 4. Some results of Weiss for silver films and of Ingersoll (1910) for bulk silver, are included. The greater scatter of the points for $nk\nu$ about the curve may be attributed to the greater experimental error in the values of $nk\nu$ compared with k^2-n^2.

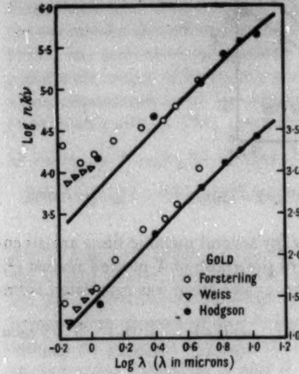


Figure 3.

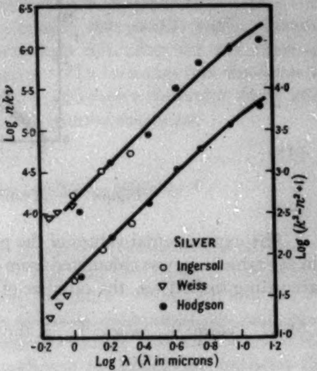


Figure 4.

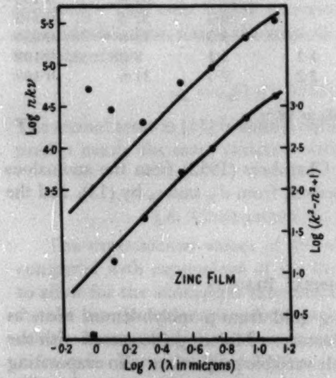


Figure 5.

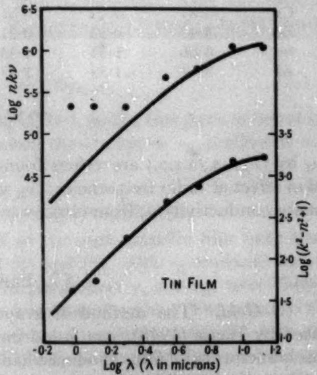


Figure 6.

(iii) *Zinc*. Zinc evaporates rapidly in vacuum at temperatures below its melting point. Films of zinc were made by evaporation from a tungsten coil. The properties of a film were measured immediately after deposition. The experimental values are compared with theoretical curves in figure 5. For λ less than 2μ the points for $nk\nu$ lie above the curve and for k^2-n^2 below the curve. This behaviour can be attributed to inter-band transitions for λ less than $2\,\mu$. The conduction band is separated by a gap of about 0·6 ev from the next energy band. The experimental value of n_0 is in good agreement with the value 0·21 deduced by Wilson (1936) from the zone structure of zinc.

(iv) *Tin.* Tin can be conveniently evaporated from a molybdenum boat. Visually opaque films of tin had cloudy surfaces of low visual reflectivity. The cloudiness increased with the thickness of the film. This may be an effect of micro-crystalline structure in tin films, which becomes coarser as the film thickness is increased. The experimental values for a tin film are compared with theoretical curves in figure 6. This film was visually feebly transmitting and had little surface cloudiness. For this film $\nu_0 = 6{\cdot}66 \times 10^4$ cm^{-1} and $\nu_R' = 1{\cdot}42 \times 10^3$ cm^{-1}, while for a thick, cloudy film of tin $\nu_0 = 5{\cdot}75 \times 10^4$ cm^{-1} and $\nu_R' = 1{\cdot}16 \times 10^3$ cm^{-1}. The variation of nk and $k^2 - n^2$ with wavelength was similar however, for all films of tin.

The experimental points for $k^2 - n^2$ agree with the curve over the whole range $\lambda = 1\,\mu$ to $13\,\mu$. Since the short wavelength part of the curve is determined almost entirely by ν_0, this indicates that the effective number of free electrons N_0 is constant for $\lambda = 1\,\mu$ to $13\,\mu$. The experimental points for $nk\nu$ lie above the curve for λ less than $4\,\mu$. An inter-band transition for λ less than $4\,\mu$ would reduce the values of $k^2 - n^2$ below the free electron curve. It seems that the infra-red properties of tin are determined by the electrons in the conduction band. The variation with wavelength could be explained by a variation in ν_R which is related to the relaxation time τ by (7).

(v) *Aluminium.* Aluminium was evaporated from a tungsten filament to make a film for infra-red measurement. The molten aluminium alloys with tungsten, so the film possibly contained some tungsten. The values of the infra-red constants of this film should be useful for practical applications in which evaporated aluminium films are used. The infra-red constants were measured immediately after deposition, to minimize the effect of oxidation. The experimental values are compared with theoretical curves in figure 7. The points show small but systematic deviations from the curves.

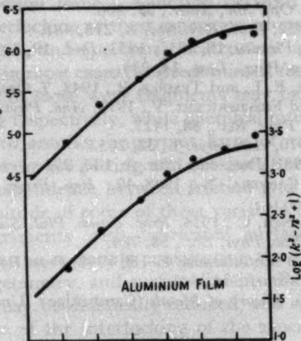


Figure 7.

(vi) *Copper*. Copper was evaporated from an alundum crucible to make a film for infra-red measurement. The infra-red constants were measured immediately after deposition. The agreement between experimental points and theoretical curves was very similar to that obtained for silver.

Several films of gold, silver and copper were prepared at various times and their infra-red constants measured. Films of the same metal were prepared by a similar evaporating procedure each time. The values of ν_0 for the same metal show variations up to $\pm 5\%$, while for ν_R the variations were up to $\pm 20\%$. For aluminium and zinc, only one film has been measured. All the metals investigated, with the probable exception of gold, rapidly form a surface layer of oxide on exposure to air. This initial layer of oxide is usually between 20 Å and 100 Å thick (Mott and Cabrera 1949). Approximate calculations have been made of the effect of a thin surface layer of oxide on the apparent infra-red constants. These calculations indicate that an oxide layer of optical thickness less than $\lambda/100$ will not affect the apparent values of n and k by more than a few per cent. In future measurements it would be desirable to find the thickness and refractive index of the oxide layer, so that its effect can be eliminated.

ACKNOWLEDGMENTS

I should like to express my thanks to Professor Redman for permission to do this work at the Cambridge Observatories, to Dr. Blackwell for his help and guidance as supervisor of the research, and to Dr. Fellgett for the design of a thermocouple amplifier. I am also indebted to the electors of the Isaac Newton Studentship for a studentship (1951–53), and to the Department of Scientific and Industrial Research for a maintenance grant (1950–51).

REFERENCES

BUECHE, F., 1948, *J. Opt. Soc. Amer.*, **38,** 806.
CHAMBERS, R. G., 1952, *Proc. Roy. Soc.* A, **215,** 481.
DINGLE, R. B., 1952, *Physica*, **18,** 985 ; 1953, *Ibid.*, **19,** 729.
DRUDE, P., 1904, *Ann. Phys., Lpz.*, **14,** 936.
ELLIOT, A., AMBROSE, E. J., and TEMPLE, R., 1948, *J. Opt. Soc. Amer.*, **38,** 212.
FORSTERLING, K., and FREEDERICKSZ, V., 1913, *Ann. Phys., Lpz.*, **40,** 201.
HOLSTEIN, T., 1952, *Phys. Rev.*, **88,** 1427.
INGERSOLL, L. R., 1910, *Astrophys. J.*, **32,** 265.
KRONIG, R. DE L., 1931, *Proc. Roy. Soc.* A, **133,** 255.
MOTT, N. F., and CABRERA, N., 1948–49, *Rep. Progr. Phys.*, **12,** 163 (London: Physical Society).
MOTT, N. F., and ZENER, C., 1934, *Proc. Camb. Phil. Soc.*, **30,** 249.
PRICE, D. J., 1946, *Proc. Phys. Soc.*, **58,** 704.
REUTER, G. E. H., and SONDHEIMER, E. H., 1948, *Proc. Roy. Soc.* A, **195,** 336.
WEISS, K., 1948, *Z. Naturf.*, **3A,** 143.
WILSON, A. H., 1936, *Theory of Metals* (Cambridge : University Press).

The Relation between Friction and Wear for Boundary-Lubricated Surfaces

By E. RABINOWICZ

Lubrication Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

MS. received 2nd March 1955

Abstract. A relationship between friction and mechanical wear for lubricated surfaces is derived in the form $Z^{2/3}=k(f-f_1)$ where Z is the wear per unit sliding distance, f the friction coefficient, and k and f_1 are constants. Experiments are described in which the metal transfer and the loose wear were measured, and satisfactory agreement with the derived equation obtained.

---

§ 1. Introduction

When contacting surfaces are rubbed together, two phenomena, namely friction and wear, invariably accompany the sliding process. In a general way it may be said that high friction and high wear go together, as do low friction and low wear, and much research has been carried out in an attempt to find a mathematical relationship between them. However, the friction and wear processes are known to be very complex, and it is not surprising that no simple relationship, certainly not a direct proportionality, has been found.

A number of factors are responsible for resistance to motion, and their sum constitutes the total friction force. Adhesion between the surface molecules of the two materials, interlocking of their asperities, ploughing of the bulk of one surface through the other, and (perhaps) electrostatic attraction all play their part, though adhesion is in most cases the predominant factor. Wear in the form of material transfer and of loose wear fragments may be produced by adhesion and asperity interlocking respectively, while chemical interaction of the lubricant with the surface leads to another important form of wear, namely corrosion. Other known forms of wear are pitting and erosion (Burwell and Strang 1952). It is clear that no correspondence between friction and wear is to be expected unless the relative magnitude of some of these variables can be diminished.

To carry out experiments with lubricated surfaces, it was decided to standardize in each case the materials constituting the sliding surfaces and the load, speed of sliding, geometry, and surface finish. By using hard metals, the influence of ploughing may be drastically reduced, and smooth surfaces serve to diminish the importance of the interlocking of the asperities. It then remains to carry out tests using a variety of lubricants, and to measure friction and wear in each case.

§ 2. Theoretical Relationship

A theoretical relationship between friction and wear may readily be derived by assuming that all the friction is due to adhesion between the metals or resistance to shear of the lubricant, while all the wear at the metallic junctions is in the form of welded metal fragments. Using the concept of Bowden, Gregory and Tabor (1945) of the lubricant partially separating the sliding surfaces while metallic contact is maintained over the remainder, we may readily derive the expression

$$f=\alpha f_m+(1-\alpha)f_l \qquad \ldots\ldots(1)$$

where f is the coefficient of friction observed, f_m and f_l the coefficients for unlubricated and perfectly lubricated surfaces respectively, and α the proportion of the surfaces that has remained unlubricated. The validity of this equation has been examined in earlier papers which have shown experimentally that f_l has a definite value independent of the nature of the contacting surfaces (Rabinowicz 1954) and that α occurs as a linear function in the equation connecting f, f_l and f_m (Rabinowicz 1955).

An analogous relation may be written for the metal transfer Z of the lubricated surfaces as a function of the metal transfer for unlubricated surfaces Z_m, but in this case only one term need be considered, since Z_l, the metal transfer between perfectly lubricated surfaces, is presumably zero. Hence we may write

$$Z=\beta Z_m \qquad \ldots\ldots(2)$$

where β represents the reduction in wear in the presence of the lubricant.

Furthermore, a relation between α and β may be derived by considering that both refer to phenomena associated with the same surface conditions and that α is an area term and β is essentially a volume term (Rabinowicz and Tabor 1951). Hence we assume

$$\alpha=\beta^{2/3} \qquad \ldots\ldots(3)$$

and combining equations (1), (2), (3) we obtain

$$\frac{Z}{Z_m}=\left(\frac{f-f_l}{f_m-f_l}\right)^{3/2} \qquad \ldots\ldots(4)$$

or

$$Z^{2/3}=k(f-f_l) \qquad \ldots\ldots(5)$$

where $k=Z_m^{2/3}/(f_m-f_l)$. This equation states that the 2/3 power of the metal transfer is proportional to the difference between the friction coefficient and some constant f_l, and may be applied to advantage when using a series of lubricants having a similar chemical structure, i.e. essentially the same value of f_l.

§ 3. RESULTS OF METAL TRANSFER EXPERIMENTS

Tests to check this law in the case of metal transfer were carried out using a number of long-chain compounds, copper palmitate, ferric palmitate, palmitic acid, cetyl alcohol, octadecyl alcohol, and cetane. Since all these have a similar chemical structure differing only in the end group, they should have the same value of f_l but they differ widely in their value of α. Data are included of tests at temperatures varying from 20 to 160°C, since the effect of a rise in temperature is to raise α while keeping f_l nearly constant. The experimental technique was that described by Rabinowicz and Tabor (1951), using a load of 2000 g, a speed of 0·01 cm sec^{-1}, and a geometry of a radioactive copper hemispherically ended rider (diameter 6·2 mm) on a flat copper surface. The metal transfer Z was calculated from the autoradiographs of the flat surface in units of g cm^{-1} of track, and each reading corresponds to a sliding distance of about 0·3 cm. The measurements of metal transfer have a probable uncertainty of about 40%. The friction coefficient was measured to ±0·005.

Since the readings cover such a wide range of Z with many points at low values of Z, it is expedient to plot the values logarithmically, i.e. to plot log Z as a function of $\log(f-f_1)$. For this purpose f_1 must be known, and a value at room temperature of 0·045 was estimated from previous work (Rabinowicz 1954) and used in our calculations. It is known that f_1 is to some extent a function of temperature, and the value of 0·045 was used for the range 20–60°c, while a value of 0·035 was adopted for the range 60–160°c. In any case, the exact magnitude of the assumed value of f_1 is of little practical importance except for very low values of f.

The results are shown plotted in figure 1. Although the 250 points show considerable scatter, they seem to lie symmetrically with respect to the straight line of slope 2/3, and are thus compatible with equation (5). At high values of $f-f_1$, above about 0·7, systematic discrepancies from the straight-line relationship become noticeable, and this behaviour may be explained by the 'snowballing' mechanism (Rabinowicz 1951), as a result of which the geometry of the sliding process is altered and equation (3) no longer applies.

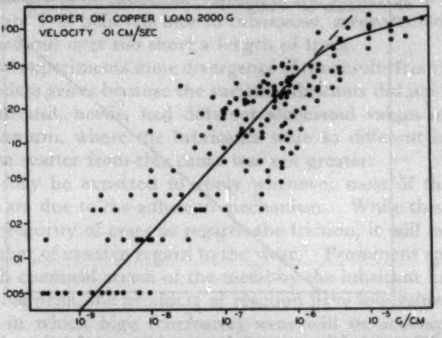


Figure 1. The friction due to metallic adhesion plotted as a function of the metal transfer, both on a logarithmic scale. The straight line has a slope of 2/3.

§ 4. Results of Loose Wear Experiments

In a previous paper it has been postulated that loose wear fragments which accompany the metal transfer process arise mainly from the filing action of the roughnesses of one surface against the other (Rabinowicz 1953), though Feng (1952) has considered a different mechanism in some detail. It seems reasonable to assume that in the presence of lubricants both the loose wear fragments and the transferred fragments will be reduced in the same proportion, so that the loose fragments will also obey equation (5). Wear tests to confirm this assumption have been carried out using the metal titanium and a wide range of lubricants.

Tests were carried out on a $\frac{1}{4}$ in. hemispherically ended rider of titanium producing a circular track of 5 cm diameter on a titanium flat. Tests were continued for about 10 000 cm at a speed of 5 cm sec^{-1}, and an average friction value was calculated from a continuous recording of the friction force. The titanium flat and rider were weighed before and after each run, and the loss in weight as result of the test was calculated. The lubricants used included mineral oils, fatty acids, polyalkylene glycols, sugar solutions, silicones, fluorocarbons, and halogenated hydrocarbons. Figure 2 shows a plot of $Z^{2/3}$ as a function of f, an arbitrary scale for $Z^{2/3}$ being used. A value of f_1 of 0·10 for these conditions was estimated from previous work with a long-chain hydrocarbon derivative (Rabinowicz 1954), and the straight line was drawn through this point. It will be seen that a very satisfactory fit of the experimental points is obtained.

As a matter of interest it was decided to recalculate the data of Neely (1937), who, using a geometry of three cylinders with flanged rims sliding on a cylinder with two flanged rails, searched for, but failed to find, a direct proportionality between F, the total friction force, and Z, the wear, his tests being carried out over a range of loads. Figure 3 shows his data obtained at one load, and the points, while showing large scattering, are consistent with our postulated relationship.

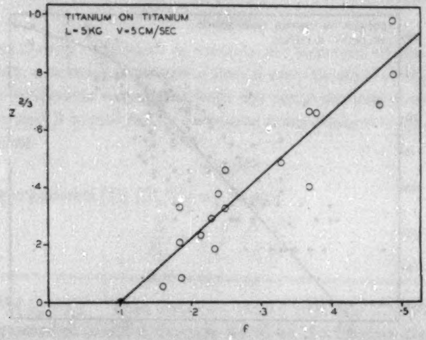


Figure 2. The 2/3 power of the wear rate (on an arbitrary scale) plotted as a function of the friction coefficient.

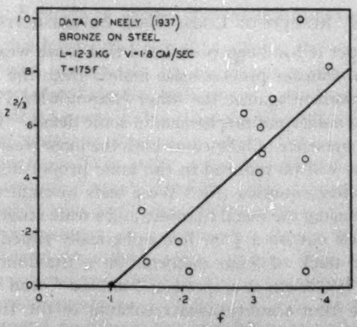


Figure 3. The 2/3 power of the wear rate (on an arbitrary scale) plotted as a function of the friction coefficient.

§ 5. DISCUSSION

The results of three different experiments have been shown to obey the 2/3 power relationship between the total wear and the metallic part of the friction force, and this provides experimental support for the theoretical treatment given earlier.

It should be emphasized that this law may be expected to apply only as long as the original wear mechanism remains the same: namely, as long as during sliding junctions are made and broken with the occasional formation of a fragment and as long as the effect of the lubricant is to change the fraction of each junction that is in metallic contact, but in other respects to leave the sliding process unchanged.

It must also be noted that the wear law is statistical in nature, since the friction force is continuous and the producton of wear fragments essentially non-continuous, and hence short-term disagreements between the postulated and observed wear rates are to be expected. Much of the scattering of results for the test shown in figure 1 is believed due to this cause, measurement of the wear having been carried out over too short a length of track.

In the other two experiments some divergence of the results from the postulated relationship doubtless arises because the various lubricants did not have a similar chemical structure and, hence, had different associated values of f_1. In fact, in the case of titanium, where the lubricants were so different in nature, it is surprising that the scatter from this cause was not greater.

Equation (5) may be expected to apply whenever most of the friction and most of the wear are due to the adhesion mechanism. While this appears to be true in the great majority of cases as regards the friction, it will not be true in a considerable number of cases in regard to the wear. Prominent among the latter are cases in which chemical attack of the metal by the lubricant has taken place. If, as frequently happens, the products of reaction have low values of f_1, we will have a situation in which high (corrosive) wear will be accompanied by low friction, and conversely (Larsen and Perry 1950). In these cases equation (5) still applies for the mechanical component of the wear, the latter being but a small fraction of the total wear.

The mathematical relationship of equation (5) shows clearly how small changes in friction can produce large changes in wear. Thus, assuming a value of f_1 of 0·045, we find that, if the friction is changed from 0·05 to 0·10, i.e. increased by a factor of 2, the corresponding increase of wear will be by a factor of 37. Furthermore, two lubricants giving the same friction coefficient of 0·05 may produce quite different amounts of wear, although their values of f_1 differ only slightly: e.g. if f_1 is 0·048 and 0·042 in the two cases, the wear produced will vary by a factor of 8. At large values of f we find, on the other hand, a better correlation between friction and wear, since small differences in f_1 will produce much smaller differences in the rates of wear.

It is of interest to note that divergences from the simple relationship of equation (5) occur at high wear rates, as shown in figure 1, and that the divergences are such that the wear obtained is larger than that postulated. Burwell and Strang (1952), who investigated the influence of load and contact area on the wear, also found that divergences were obtained at high wear rates, and again the amount of wear was greater than expected. It appears that severe wear conditions are governed by laws about which we as yet know very little, though a considerable degree of progress has been made in formulating the quantitative laws of mild wear.

ACKNOWLEDGMENTS

The author is indebted to Professor B. G. Rightmire for many helpful discussions and to Mr. E. P. Kingsbury for carrying out the wear experiments with titanium.

REFERENCES

BOWDEN, F. P., GREGORY, J. N., and TABOR, D., 1945, *Nature, Lond.*, **156,** 97.
BURWELL, J. T., and STRANG, C. D., 1952, *J. Appl. Phys.*, **23,** 18.
FENG, I-MING, 1952, *J. Appl. Phys.*, **23,** 1011.
LARSEN, R. G., and PERRY, G. L., 1950, *Mechanical Wear*, ASM, 73.
NEELY, G. L., 1937, *Proc. Disc. Lubrication and Lubricants*, Vol. II, Instn Mech. Engrs, 378.
RABINOWICZ, E., 1951, *J. Appl. Phys.*, **22,** 1373; 1953, *Ibid.*, **24,** 367; 1954, *Lubric. Engng*, **10,** 205; 1955, *Proc. 2nd U.S. Congress of Applied Mechanics, Amer. Soc. Mech. Engrs*, 595.
RABINOWICZ, E., and TABOR, D., 1951, *Proc. Roy. Soc.* A, **208,** 455.

Deformation Properties of Friction Junctions

BY J. A. GREENWOOD AND D. TABOR
Research Laboratory for the Physics and Chemistry of Surfaces,
Department of Physical Chemistry, Cambridge

MS. received 17*th March* 1955 *and in amended form* 11*th May* 1955

Abstract. This paper describes an experimental investigation of the mechanism of sliding friction by shearing large-scale models of friction junctions under conditions where no normal displacements are allowed. This condition, as A. P. Green first pointed out, must hold in steady sliding. Experiments have been performed simulating the cases of very strong adhesion, weak adhesion, and lubricated sliding with very little adhesion. The results are in good qualitative agreement with the theoretical solutions obtained by Green using the theory of plasticity. For very strong adhesion the models yield average values of μ in the range 3–10. Coefficients of friction of the order $\mu = 1$ are only obtained for weak adhesion. This suggests that the value observed for clean metals sliding in air ($\mu = 1$) results from the presence of oxide or contaminant films. Results for lubricated junctions, where adhesion is negligible, suggest that in lubricated sliding, plastic deformation of the sliding metals may contribute an appreciable part of the observed friction even when there is no actual penetration of the lubricant film.

§ 1. INTRODUCTION

According to the adhesion theory of friction, when one metal slides on another, contact at any instant occurs only at a small number of minute asperities. Because the total area of real contact is very small the high stresses developed at these regions are sufficient to cause the metals to flow plastically giving rise to strong adhesion between them by a process analogous to cold welding. The friction force F is the force required to shear these welded junctions. If p is the mean normal stress over the regions of contact in sliding, and s the mean shear stress, then the load $W = Ap$, and $F = As$, where A is the real area of contact. In the original formulation of the theory p and s were considered to be constant strength properties of the material, so that $A = W/p$, i.e. the true area of contact is proportional to the load, and independent of the apparent area of contact, while $F = Ws/p$, i.e. the friction force obeys the two basic laws of friction: it is proportional to the load and independent of the apparent area of contact. This simple mechanism has also proved valuable in explaining many other frictional phenomena.

If the simple theory is extended too far, difficulties arise. At first sight it seems reasonable to identify p with the yield stress in static loading, p_m. But it can be shown that if k is the critical shear stress at which plastic flow begins, then for an ideal plastic material, $p_m \sim 6k$ (Tabor 1951). If it is also assumed that $s \sim k$, then the coefficient of friction $\mu = F/W = s/p \sim k/6k = 0{\cdot}17$ for an ideal plastic material, and may be expected to have a similar value for all metals. In practice it is found that for most clean metals in air μ is near 1, so that at least one of these assumptions must be wrong.

In fact it is clear from modern plasticity theory that plastic deformation takes place under the combined effect of the normal and tangential stresses so that the values of these at which yielding takes place are not constant properties of the material, but are interdependent and may vary during the life of the junction. A junction of given shape has a yield criterion involving both p and s. This was first pointed out by McFarlane and Tabor (1950), who found experimentally using indium, a very soft, ductile metal, that the yield criterion was $p^2+\alpha s^2=p_m^2$ where α had the value 3·3.

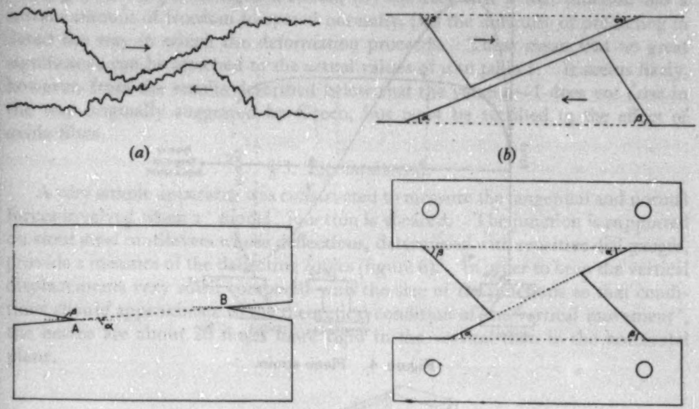
The consequences of this are two-fold. In static loading with no tangential force the junctions are already at the yield point under the normal load alone. As a result, the smallest tangential force is sufficient to upset the equilibrium, and some motion occurs. This will involve both a small tangential displacement and a sinking together of the surfaces, so that the junctions increase in size and number. This constitutes an increase in the area of contact. In the new state of equilibrium the normal load is supported over an increased area and so produces a decreased normal stress. Consequently the tangential stress must be increased further before sliding will continue. Both these consequences have been observed experimentally; the initiation of tangential movement for the smallest tangential force by Cocks (1954) and by Eisner (1954) and the continuous growth of the junctions under these conditions by McFarlane and Tabor (1950), and by Parker and Hatch (1950).

Although this application of the modern theory of plasticity explains junction growth during the initial stages of microdisplacement it is insufficient to explain why, with most metals in air, the process comes to an end and steady state sliding generally takes place when μ has a value near unity. Green was the first to point out that the characteristic feature of *steady* sliding which had been overlooked is that the surfaces move parallel to themselves, each junction shearing without undergoing any normal motion. This is not an assumption but is implicit in the conception of a steady state and it is this condition which determines the relationship between p and s. The purpose of this paper is to describe an experimental examination of the shearing of a simulated friction junction, under conditions where no normal motion is permitted.

§ 2. THEORETICAL JUNCTION MODEL

In his theoretical treatment Green (1955) considers a simplified form of asperity shown in figure 1 (*a*) and (*b*). In the first instance he assumes strong adhesion across the interface AB. He postulates no normal motion during shearing of the junction (i.e. he neglects small scale motion that may be possible on account of the elasticity of the metals) and in this way obtains a soluble problem in plasticity. The force to shear the junction depends on the geometry of the junction and this has been solved for two extreme cases : (i) plane strain where the junction is very thick in the direction normal to the plane of the paper, (ii) plane stress where the junction thickness is uniform and very small. His solution gives only the values of the forces needed to deform certain simple shapes of junction. Some typical results for the plane strain and plane stress solutions are given in figures 2 and 3 respectively. It is evident that there is an intrinsic inadequacy in the plane strain solution for very small values of α and β: for in the limiting case of a long shallow junction where $\alpha\simeq\beta\simeq 0$ (figure 1 (*c*)) one obvious solution is pure shear across the neck AB, which can take place for any value of the normal stress p

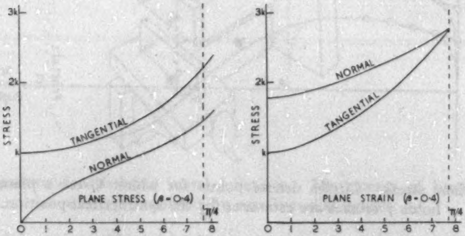
(including $p=0$) so that the normal force W is indeterminate. Its value can only be found by abandoning the model of a plastic-rigid material and taking into account the elastic deformation. This suggests that the values of W predicted for small but non-zero values of α and β also are likely to be critically dependent on the plastic-rigid assumption.



(*a*) (*b*)

(*c*) (*d*)

Figure 1.



Figures 2 and 3. Forces required to deform a junction of given shape.

In the next stage of his analysis Green constructed some plasticine models from which he was able to form a general picture of the way in which deformation proceeds during the shearing process. He observed that α decreased slowly whilst β increased fairly rapidly to values well beyond 90°. In this way he was able to make an estimate of the theoretical forces throughout the deformation of a junction. Typical results for the plane strain and plane stress solutions are given in figures 4 and 5 respectively. The plane strain solution shows three main features : (i) a fairly constant value of the tangential force F until the junction fractures, (ii) a rise in the normal force W during the initial stage of deformation and a decrease to zero when the junction reaches a symmetrical shape, (iii) a subsequent stage in which the junction is in tension (i.e. W is negative) until fracture occurs. The plane stress solution is valid over such a small range of β that it is more difficult to predict how the forces will vary. However it seems likely that the plane stress solution will show the same features qualitatively as the plane strain solution, though the relative values of F/W will be very different.

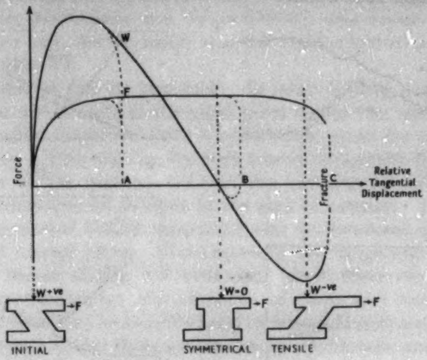

W
F
Force
A
B
C
Fracture
Relative
Tangential
Displacement
W+ve
→F
W=0
→F
W-ve
→F
INITIAL
SYMMETRICAL
TENSILE

Figure 4. Plane strain.

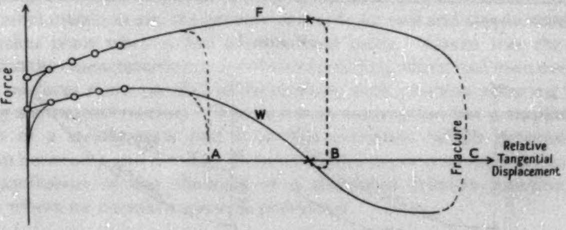

F
Force
W
A
B
C
Fracture
Relative
Tangential
Displacement

Figure 5. Plane stress. Circles denote points for which Green's plane stress solution holds ; crosses are estimates for the symmetrical position.

If two metals are in contact over a large number of junctions, such that some are being formed whilst others are being sheared, the average value of the forces over all the junctions at any instant will be equal to the average value of the forces involved in a single junction over its complete life cycle. Using the results of figures 4 and 5 it is thus possible to estimate the ' coefficient of friction ' of the junction, defined as the ratio of the average tangential force to the average normal force (see table 1). This value clearly depends on the point of rupture of the junction, and in this way on whether the material is brittle, normal, or ductile.

Table 1. Theoretical Coefficients of Friction

Mode of deformation	Plane strain	Plane stress
Very brittle solid (fracture at A)	0·4	1·5
Normal solid (fracture at B)	0·8	3·0
Very ductile solid (fracture at C)	1·0	4·0

Green originally suggested that the value of μ of the order of 1 observed in practice for clean metals in air corresponded to the behaviour of a moderately ductile junction under plane strain conditions. But as Green points out in a later paper, the behaviour of a real junction will be complicated by three factors which will be difficult to allow for numerically: (i) the inadequacy of the assumption of a plastic-rigid material, (ii) the fact that a real junction has a certain amount of freedom to spread normally, (iii) the difficulty of predicting in detail the way in which the deformation proceeds. These mean that no great significance can be attached to the actual values of μ in table 1. It seems likely, however, from the results described below that the value $\mu \sim 1$ does not arise in the way originally suggested by Green, but must be ascribed to the effect of oxide films.

§ 3. EXPERIMENTAL

A very simple apparatus was constructed to measure the tangential and normal forces involved when a ' model ' junction is sheared. The junction is supported on stout steel cantilevers whose deflections, determined with sensitive dial gauges, provide a measure of the deflecting forces (figure 6). In order to keep the vertical displacements very small compared with the size of the junctions so that conditions should approximate to the theoretical condition of ' no vertical movement ', the beams are about 20 times more rigid in the vertical than in the horizontal plane.

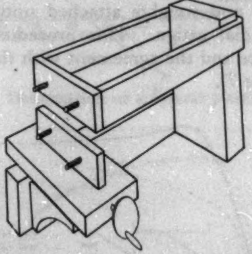


Figure 6. Apparatus for shearing metal junctions. Movements of beam upwards or sideways recorded by sensitive dial gauges.

The model junctions (figure 1 (*d*)) used in the first stages of the work were fabricated from a single piece of material. This provides the closest approximation to the theoretical model which assumes very strong interfacial adhesion. Materials used were annealed aluminium, lead, indium and plasticine.

§ 4. RESULTS

4.1. *The Initial Forces for Junctions of Different Shapes and Thicknesses*

Experiments were carried out on model junctions to see how far the observed initial forces agreed with the predictions of the theory (figures 2 and 3). It was found that a finite shear was necessary before representative values of normal and tangential forces were obtained. This is probably because the yield point is not sharply defined. The tangential forces agreed reasonably well with theory, but the normal forces were always considerably less than the theoretical values.

The first junctions used were thin compared with their length so that conditions would approximate to plane stress. Using plasticine it was possible to measure the forces on very thick junctions (thickness six times the length) so that conditions were closer to those of plane strain. This produced no appreciable change in the normal stress. The normal forces remained small and were of the same order as those predicted by the plane stress model. In particular it was found that for long shallow symmetrical junctions ($\alpha \simeq \beta \simeq 0$) the normal force was negligibly small. Similar results were observed using lead junctions (thickness twice the length). It was concluded that under the conditions of the investigation the plane strain solution found little experimental justification.

With aluminium (on account of its very much greater strength) it was not possible to use thick junctions. Therefore most of the subsequent work was carried out on metal junctions cut out of thin sheets for which the plane stress theory might be expected to hold. With these junctions the tangential forces were in reasonably close qualitative and quantitative agreement with the theoretical values. The tangential forces were higher when α was increased as the theory predicts (see figure 2). The normal forces also increased with α (see figure 2) but were 30–40% less than the theoretical values. This means that the ratio W/F is also 30–40% less than the values predicted by the plane stress theory (figure 7). It was considered that this might be due to the small amount of vertical yielding in the cantilevers which would tend to reduce the normal force. To overcome this a lever of the second kind was constructed over the main cantilever arms and weights attached until no net movement was observed in the vertical dial gauge. This procedure led to an appreciable increase in the normal force and the agreement with the plane stress theory was more satisfactory (see figure 7).

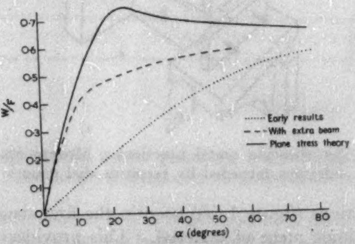


Figure 7. Forces required to deform junctions of different shapes.

These preliminary measurements show that the initial solutions for junctions under plane stress conditions are satisfied experimentally. Slight vertical yielding causes an appreciable drop in the normal force. Large increases in the thickness do not appear to 'shift' the results towards the plane strain solution. This is due in part to the materials from which the thick junctions were made. Plasticine has to undergo a large strain (~10%) before it gets round the bend of the stress–strain curve and deforms in a truly plastic manner. Lead also has a badly defined yield point, and in fact it is only for work-hardened metals or for very large strains such as those involved in strip-rolling and other technological processes that the plastic-rigid theory gives accurate results. There is also the consideration that the vertical force is a constraint and does no work. Such forces tend to take almost arbitrary values, i.e., they depend critically on minor factors. This is probably the main reason for the difference between the value of the normal force predicted by the plane stress theory and that predicted by the plane strain theory. On the other hand the tangential force, which does do work, is predicted to have almost the same value on either theory. This variability of the normal force suggests that a very small amount of yielding, either vertical or lateral, could completely alter its value. It is plain that the same will apply to the actual friction junctions formed between sliding surfaces.

4.2. *Deformation of Typical Model Junctions*

The life cycle of a typical plasticine junction (before the additional lever was fitted) is given in figure 8. It is seen that it reproduces all the main features of the theoretical solution. The shearing force is roughly constant throughout the deformation whilst the normal force rises to a maximum, falls to zero when the junction shape is symmetrical, and then becomes negative when the junction is subjected to tension. Substantially the same results were obtained for thin junctions as for very thick ones, the behaviour in all cases being fairly near the theoretical results derived for deformation under plane stress. With plasticine the negative loop may have an area larger than the positive so that the overall

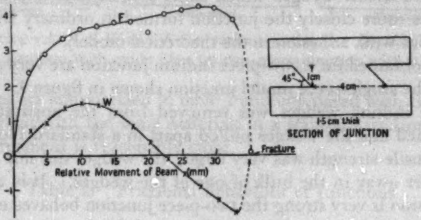


Figure 8. Deformation of a typical plasticine junction.

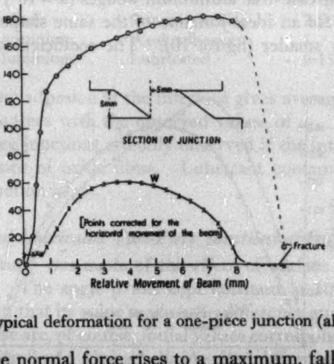


Figure 9. Typical deformation for a one-piece junction (aluminium).

average value of W over the life-cycle is negative. This means that if plasticine surfaces are slid together a finite friction may be observed even if the normal load is reduced to zero or even made tensile.

With lead a similar negative loop was obtained before the junction fractured. The overall average value of W was positive but very small so that the coefficient of friction was very large; it depended on the shape of the junction but in general lay between 3 and 10.

With aluminium no negative portion of the normal force curve was observed. A typical result obtained when the additional lever was applied to prevent vertical movement is given in figure 9. For a junction of this shape $\mu \simeq 3$–4.

4.3. *Deformation of Typical Two-Piece Junctions*

In the experimental study so far described, the junctions have all been cut from a single piece of material to simulate the case of very strong adhesion discussed in the theoretical model. It is of interest to compare this with the behaviour of two model wedges which come into contact and form a junction which resembles more closely the junction formed in ordinary frictional experiments (the case of weak adhesion in the theoretical model).

The results obtained for a two-piece indium junction are very similar to those observed with the single-piece model junction shown in figure 9. In one experiment a pair of indium wedges was removed from the apparatus after being partially deformed and the wedges pulled apart in a standard Hounsfield tensometer. The tensile strength was very large; the wedges did not separate at the interface but tore away in the bulk of one of the wedges. It is clear that when interfacial adhesion is very strong the two-piece junction behaves essentially like a single-piece model junction. The coefficient of friction (for $\alpha = 10°$) is again about 2·5.

With two clean grease-free aluminium wedges ($\alpha = 10°$) the normal forces are almost the same as for an ideal junction of the same shape, while the tangential force is appreciably smaller (figure 10). The coefficient of friction is about 1.

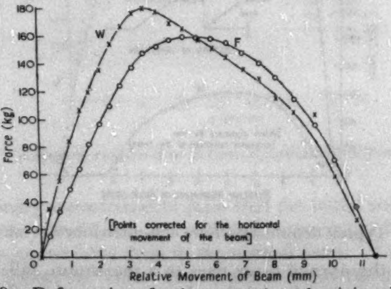


Figure 10. Deformation of a clean two-piece aluminium junction.

Observation showed that adhesion between the wedges was small; some yielding appeared to take place in the vertical direction and there was slight slipping of one wedge over the face of the other during the deformation process.

The only contamination on the aluminium was its oxide film and it would seem that this reduces the strength of adhesion at the interface and so decreases the overall value of μ. These effects are all in accordance with Green's theoretical results for weak adhesion.

If a lubricant such as a metallic soap is deliberately added to the aluminium wedges these trends are more pronounced and the coefficient of friction (for $\alpha = 10^\circ$) is about 0·15. The main characteristics of the three types of junction are shown in figures 11 (*a*), 11 (*b*) and 11 (*c*).

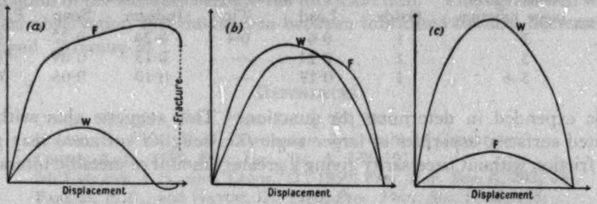


Figure 11. Relative values of W and F for various junctions. (*a*) single-piece junction, (*b*) two-piece junction with oxide film, (*c*) two-piece junction with lubricant film.

These results are summarized in table 2 and compared with the coefficient of friction μ_s observed with the same materials in a standard frictional experiment. The general agreement is reasonably close.

Table 2. Frictional Behaviour of Junctions

Type of junction ($\alpha = 10^\circ$)	Material	Surface condition	μ	μ_s
One-piece model	Any	Perfect adhesion	3	steady sliding not observed
Two-piece	Indium	Strong adhesion	2·5	2
Two-piece	Aluminium	Weak adhesion	1	1
Two-piece	Aluminium	Lubricated	0·15	0·05

It is seen that strong adhesion at the interface gives average values of μ of the order of 2–3 and this agrees with the observed values of μ_s. Values of μ of the order of 1 for two-piece junctions are only observed if the interfacial adhesion is reduced by the presence of oxide films. Lubricant contamination produces a very much larger reduction in μ.

4.4. *Behaviour of Lubricated Two-Piece Junctions of Different Shapes*

A more detailed study was made of the effect of surface films on two-piece aluminium junctions. The angle β was kept constant at 30° after preliminary experiments had shown that its value made little difference and the angle α varied from $3\frac{1}{2}^\circ$ to 14°. These are, of course, initial values corresponding to the shape of typical surface asperities and change during the deformation process. The lubricants used were lead, tin and indium foils, and thin films of metallic soap and of P.T.F.E. The results are summarized in table 3.

It is seen that oxide films reduce μ from its value for ideal junctions of 3–5 to about $\mu = 1$ and that lubricant films reduce the value of μ by a further factor. The biggest reductions are observed with soap and with P.T.F.E. films.

For well lubricated surfaces μ decreases with decreasing α. With the P.T.F.E. films the electrical resistance across the junction was measured and it was found that the resistance remained infinite throughout the greater part of the deformation of the junction. This means that the increase in μ with α is not primarily due to increased metallic interaction. It arises from the fact that more plastic work is expended in deforming the junction. This suggests that with well lubricated surfaces, asperities of larger angle (i.e. rougher surfaces) may give a higher friction without necessarily giving a greater amount of metallic interaction.

Table 3. Effect of Lubricant Films on Behaviour of two-piece Aluminium Junctions—values of μ

α(°)	Ideal single-piece junction	Two-piece junctions with films of					
		Oxide	Lead	Tin	Indium	Soap	P.T.F.E.
14	2	1	0·6	0·4	0·26	0·22	0·23
7	3	1	0·24	—	0·15	0·09	0·19
$3\frac{1}{2}$	5–6	1	0·17	—	0·10	0·05	0·10

§ 5. FRICTION EXPERIMENTS

Some experiments were made to detect any dependence of friction on surface roughness. Lead surfaces were prepared, covered by a set of parallel ridges which had one face sloping at 30° and the other at 60°. These were covered with oleic acid and the friction measured forwards and backwards, perpendicular to the ridges, at a load great enough for the slider to be in contact at any instant with about ten ridges. Preliminary results suggest that the friction in the direction up the 60° slopes (corresponding to $\alpha=60°$, $\beta=30°$ in the previous work) was consistently higher than the friction in the other direction ($\alpha=30°$, $\beta=60°$) by a small amount ($\sim 10\%$). This may be due to increased breakdown of the lubricant film. On the other hand, the results are also consistent with the view expressed above that in lubricated sliding the friction depends to some extent on the amount of plastic work done on the metal through the lubricant film.

§ 6. CONCLUSION

The first and most important conclusion of this experimental study is the close qualitative agreement with Green's theory of the deformation of a friction junction sheared under conditions of little or no normal motion. In addition it is found that the ratio of the tangential and normal forces averaged over the life-cycle of a single junction is close to the observed coefficient of friction in an ordinary sliding experiment. Thus two basic elements of Green's theory are qualitatively valid. Quantitatively the agreement is good for the plane stress solution but does not appear to be satisfactory for the plane strain solution.

Ideal junctions involving very strong adhesion at the interface give, for small angle of the surface asperities ($\alpha=3\frac{1}{2}°$), an average coefficient of friction of over 3, and for ductile materials values of μ greater than 10 may be attained. This is comparable with the friction observed between thoroughly clean metals *in vacuo*. The observed values of μ of the order of 1 for clean metals in air must be ascribed primarily to the effect of oxide films.

Lubricant films produce a further reduction in the average friction. For steeper asperities (larger α) the average friction is increased. This does not appear to be due to increased penetration of the film leading to increased metal–metal interaction: it arises primarily from the greater amount of plastic work expended in deforming the metal beneath the lubricant film.

Ordinary friction experiments between lubricated surfaces as distinct from experiments with 'model' junctions support this view. The friction is higher when steeper asperities are present. This suggests that in lubricated sliding the friction is due not only to the shearing of the lubricant film and of any metallic junctions formed by penetration of the film; there is also a term due to the plastic deformation of the surfaces beneath the lubricant film. This agrees well with the practical observation that the friction between lubricated surfaces decreases after prolonged 'running-in'.

REFERENCES

COCKS, M., 1954, *Proc. Phys. Soc.* B, **67**, 238.
EISNER, E., 1954, *Ph.D. Dissertation*, University of Cambridge.
GREEN, A. P., 1955, *Proc. Roy. Soc.* A, **228**, 191.
MCFARLANE, J. S., and TABOR, D., 1950, *Proc. Roy. Soc.* A, **202**, 244.
PARKER, R. C., and HATCH, D., 1950, *Proc. Phys. Soc.* B, **63**, 195.
TABOR, D., 1951, *The Hardness of Metals* (Oxford : Clarendon Press).

Effects Produced by Metal Electrodes in Studies of Electron Bombardment Conductivity

By J. C. FIRMIN and C. W. OATLEY
Engineering Laboratory, University of Cambridge

MS. received 21*st February* 1955 *and in final form* 25*th April* 1955

Abstract. In studies of electron bombardment conductivity, it has commonly been the practice to coat one face of the material under investigation with a metallic film, which can act as an electrode. If such a film is sufficiently thin the incident electrons can pass through it without serious loss of energy.

The experiments described show that, in some cases, the apparent electron bombardment conductivity is greater with thick films than with thin. They show further that this effect can be ascribed to the generation of x-rays in the film. The bearing of these results on previous measurements of electron bombardment conductivity is discussed.

---

§ 1. Introduction

When an insulating crystal is bombarded with electrons of sufficiently high energy, some of the electrons in the crystal will be raised to the conduction band and the crystal may then show appreciable conductivity. This effect is known as electron bombardment conductivity and has been studied by several people including Ehrenberg and Ansbacher (1951), Pensak (1949, 1950), McKay (1948, 1950) and Benda (1951).

Although the basic phenomenon is simple, any attempt to study it is subject to two major difficulties. The first of these arises from the fact that the bombarded crystal is an insulator. Unless steps are taken to stabilize the potential of the surface on which the electrons impinge, charges will be built up and the effective energy of the incident electrons, as well as their distribution over the surface, will be subject to considerable uncertainty. In the work of Pensak, stabilization of the surface potential was achieved by flooding the surface with electrons (from a separate source) whose energy was too low to cause appreciable bombardment conductivity, but most other investigators have coated the face of the crystal with a metallic film which acts as an electrode, but which is not thick enough to cause serious loss of energy from the incident electrons. It is with the effects produced by metallic-film electrodes of this type that the present paper is concerned.

The second major difficulty is caused by the presence in the crystal of electron traps. As a result of these, space charge is built up in the crystal and, if the incident electron current and the field applied to the crystal are both steady, the space charge may eventually neutralize the applied field so that flow of current through the crystal ceases. Two general methods of overcoming this difficulty have been employed. In the first of these the conditions are such that the primary electrons penetrate the whole of that region of the crystal through which the flow of current is measured. Thus, in the work of Pensak and of Ehrenberg and Ansbacher, the insulator was in the form of a thin polycrystalline film through which the majority of the primary electrons were able to pass. The continuous generation of hole/electron pairs throughout the film ensured rapid neutralization of any trapped particles and thus prevented the building up of excessive space charge. A somewhat similar result was achieved by Benda, who used two electrodes on the same face of the crystal and bombarded the narrow area between them.

While experiments of this kind may provide results of considerable technological importance, they are rather limited in the information they can give about the fundamental processes taking place inside the crystal. A second method of allowing for the effects of space charge has therefore been used by McKay, who worked with a relatively thick crystal and bombarded it with pulses of electron current of only a few microseconds duration. Between pulses the crystal was subjected to treatment which removed space charge so that, at the beginning of each pulse, the only field in the crystal was the externally applied one. The current through the crystal was measured with an oscillograph and the rapid build-up of space charge could readily be observed and taken into account.

A somewhat analogous method has been studied in this laboratory. The results obtained with it have proved difficult to interpret and will not be discussed in detail, but it has provided some very definite information about the large effect that can be caused by the metallic-film front electrode which is commonly used in such work, and this information is reviewed below.

§ 2. METHOD AND APPARATUS

The experiments were made on a diamond which was a nearly perfect single crystal and which was a satisfactory specimen for counting alpha particles. It was of a fairly uniform thickness of about 0·15 cm and its large faces were roughly triangular in shape. On these faces silver electrodes about 0·3 cm in diameter were evaporated, the front one being sufficiently thin to permit the passage of the primary beam of electrons without undue loss of energy.

The diamond was mounted in front of an electron gun in an evacuated vessel, so that a focused beam of electrons could be projected on to its front face. It was enclosed in a metal screening box in one side of which was a hole, 0·07 cm in diameter, for the passage of the incident electrons. The diamond with its two electrodes formed a small condenser which was connected in one arm of a sensitive a.c. bridge of the type described by Clark and Vanderlyn (1949), in which closely coupled coils are used as ratio arms. Such bridges are almost immune from the effects of stray capacitances to earth, and they permit the use of very high ratios so that small capacitances can be measured against standard condensers of convenient size. The circuit used is shown in figure 1 where C_1 shunted by r represents the specimen, C_2 is a standard condenser and r_1, r_2 and R are standard resistances. The incident electron current was returned to the cathode through a suitable meter and means were provided for applying a d.c. field to the crystal, though this was not needed in the experiments under review.

When the diamond is bombarded, two changes will occur which may be expected to alter the effective capacitance and parallel resistance of the condenser of which it forms the dielectric. Firstly, a layer of the diamond adjacent to the front electrode will be penetrated by the incident electrons and large numbers of holes and electrons will be formed in it. This region, which may be termed the plasma, will thus become conducting though its resistance will not be negligible. In consequence, the overall capacitance will increase and so also will the power factor. In the second place, during one half-cycle electrons will be drawn from the plasma into the rest of the crystal while, during the following half-cycle, holes will likewise be drawn in. The currents thus caused to flow will not, in general, be in phase with the voltage across the condenser and, for our present purpose, it is not necessary to analyse them in detail. They will cause further changes in the effective capacitance and power factor.

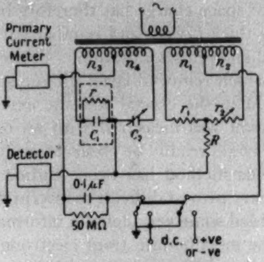


Figure 1. Bridge circuit.

It is possible to make a rough estimate of the depth to which the primary electrons penetrate and from this to calculate the change in capacity to be expected from the short-circuiting effect of the plasma. This turns out to be only about 1% of the observed change, so both of the processes described above must be operative.

§ 3. Results

For our present purpose it is sufficient to regard the change of capacitance δC or the change of parallel resistance δr, produced by bombardment as a measure of the induced conductivity, without attempting to relate either to any other parameters. In general the changes in δr were very similar to those in δC, so only the latter will be considered.

Figure 2 shows curves, for two different values of bombarding current, of δC plotted against the energy of the incident electrons. The change becomes very small below about 6 kev and it seemed probable that this value represented the mean energy lost by the electrons in passing through the silver film. Measurements were therefore made using the same crystal but with front electrodes of varying thickness, and the results are shown in figure 3. The broken lines represent attempts to extrapolate those parts of the curves which correspond to conditions when nearly all of the incident electrons pass through the electrode. At lower energies some of the electrons will still get through and we should expect the values of δC to tail off gradually. Clearly such extrapolations are somewhat arbitrary, but the values of the threshold energies obtained from them are in rough agreement with the Thomson–Whiddington law.

What is more important is that, although the values of δC are increased by reducing the thickness of the front electrodes when the energy of the incident electrons is fairly low, the reverse is true at higher energies. It seemed unlikely that this result could have been brought about by the emission of secondary electrons into the crystal from the back face of the front electrode, but to check this point some rough measurements were made with silver films deposited on a collodion substrate about 5×10^{-6} cm thick. These indicated that, with an incident energy of 10 kev and with thicknesses of silver ranging from 5×10^{-6} to $2{\cdot}5 \times 10^{-5}$ cm, the number of low-voltage (<80 ev) secondary electrons issuing from the back face did not exceed about 15% of the primary current. Furthermore, the number decreased with increasing primary energy to about 5% at 20 kev.

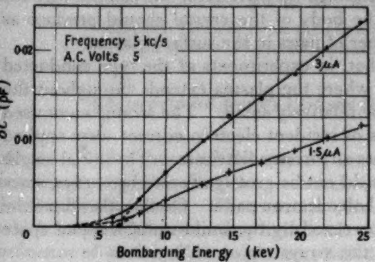


Figure 2. Variation of δC with the energy of the incident electrons.

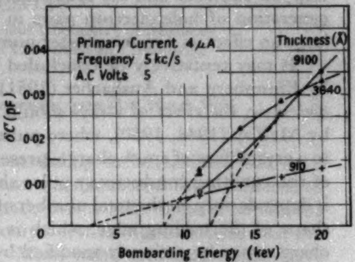


Figure 3. Variation of δC with electrode thickness.

The most probable explanation of the increased effect with thick films appeared to be the generation of x-rays in the metal, and this was confirmed as follows. A piece of tungsten foil 10^{-4} cm thick was mounted in front of, but not touching, the diamond, which was provided with its normal front electrode. This thickness of tungsten was sufficient to stop the incident electrons, whose energy did not exceed 20 kev, so that none of them reached the front electrode of the diamond. The variation of δC with incident electron current and energy was now found to be very similar to that obtained previously, except that the new values of δC were about four times as great as the old.

§ 4. DISCUSSION

The experiments described above show that the generation of x-rays in the front electrode can, under certain conditions, seriously affect the results obtained in investigations of electron bombardment conductivity and it is interesting to recall that, in the work of Benda (1951), increased conductivity was observed whenever the incident beam was allowed to strike one of the metal electrodes rather than the crystal surface between them. The extent of the part played by x-rays will depend very much on the particular experimental arrangement. Whenever the incident beam has been allowed to pass through a metal film electrode, investigators have always been careful to ensure that only a small proportion of the initial energy was lost thereby. Thus, even if the whole of this energy were converted into x-rays and if all the x-rays penetrated into the crystal, it would still be safe to conclude that the number of hole/electron pairs generated by x-rays from the metal was small in comparison with those generated directly by electrons reaching the crystal. The large effects produced by x-rays in the present experiments must therefore be ascribed to the fact that the electrons only generate pairs very near the surface, whereas the x-rays penetrate more deeply into the crystal. Since the magnitude of δC is largely controlled by the behaviour of trapped electrons and the space charge which they produce, one can see why the generation of hole/electron pairs in the body of the crystal should produce as large an effect as a much greater number of pairs in the surface plasma.

It may tentatively be concluded that, in experiments of the type conducted by Ehrenberg and Ansbacher (1951), where the plasma extends throughout the dielectric, the effect of x-rays should be relatively small. The arrangement used by McKay (1948, 1950), where pulses of incident electron current were applied to a crystal free of space charge, presents an intermediate case. The initial height of the resulting conductivity pulse should not be greatly affected by x-rays, since it depends only on the total number of hole/electron pairs. On the other hand the shape of the trailing edge, which is governed by the build-up of trapped space charge, may be seriously modified by the x-rays.

X-rays may, of course, be generated in the crystal itself but consideration of any effects which these may produce is outside the scope of the present paper.

ACKNOWLEDGMENTS

We are greatly indebted to Mr. W. F. Cotty of the Diamond Producers' Association for the loan of the diamond and for the trouble which he took to find us a suitable specimen. We are also very grateful to the Admiralty for a grant to cover the cost of the apparatus.

While the work was being carried out, one of us (J. C. F.) was in receipt of a Department of Scientific and Industrial Research maintenance grant.

REFERENCES

BENDA, H., 1951, *Ann. Phys., Lpz.*, **9**, 413.
CLARK, H. A. M., and VANDERLYN, P. B., 1949, *J. Instn Elect. Engrs*, Pt. 3, **96**, 189.
EHRENBERG, W., and ANSBACHER, F., 1951, *Proc. Phys. Soc.* A, **64**, 362.
MCKAY, K. G., 1948, *Phys. Rev.*, **74**, 1606; 1950, *Ibid.*, **77**, 816.
PENSAK, L., 1949, *Phys. Rev.*, **75**, 472 ; 1950, *Ibid.*, **79**, 171.

The Contact Resistance and Mechanical Properties of Surface Films on Metals

By R. W. WILSON[†]
Research Laboratory on the Physics and Chemistry of Surfaces,
Department of Physical Chemistry, University of Cambridge

MS. received 3rd January 1955 *and in amended form 21st March* 1955

Abstract. The nature of the contact between sliding metal surfaces has been investigated by measuring the electrical contact resistance and the coefficient of friction between them. Observations were made over a wide range of conditions which extended down to very light loads of only a few milligrammes so that the influence of surface films might be studied. Many noble and base metals have been investigated in both clean and lubricated states. The contact between certain typical pairs of dissimilar metals has also been studied.

It was found that at loads below a few grammes the oxide film on base metals prevented actual metal-to-metal contact almost entirely. At higher loads the film was disrupted, and metallic contact occurred. The change from oxide-on-oxide to metal-on-metal sliding was usually accompanied by changes in the coefficient of friction, the nature of the sliding, and the appearance of the track. This, however, was not invariable and occasionally changes in the nature of sliding occurred which were detectable only by measurements of the contact resistance. The experiments emphasize the importance of the measurement of contact resistance as a tool in the investigation of the behaviour of non-conducting surface films. The loads at which the oxide films on various metals, in particular platinum, disrupt have been compared; the manner in which the oxygen layer is attached to the metal and the physical factors controlling the break-through have been studied.

The same techniques have been used to investigate the influence of the Beilby layer, by comparing the behaviour of annealed and electropolished surfaces with that of worked ones.

The contact between lubricated metals has also been studied over a wide range of loads and the electrical contact measurements have shown that with monobasic fatty acids a change in the nature of contact occurred as the load was increased. It was found that at heavy loads the contact resistance between lubricated metals fell to a value comparable with that of clean surfaces although measurements of the coefficient of friction suggested that the area of actual metallic contact had not increased correspondingly.

These experiments have shown that the behaviour of oxide, Beilby, and lubricant films on metal surfaces follow the same pattern, and that a single general picture, based on the theory of the sliding between clean metals at heavy loads, can be given which will describe their influence on the contact and sliding between metals.

[†] Now at Thornton Research Centre, P.O. Box 1, Chester.

§ 1. INTRODUCTION

It is only under very specialized conditions (for example, in a high vacuum at high temperature) that a metal surface can exist free from any contaminating film. The contaminating films encountered on a typical metal in air may be divided into three main groups: oxide films, worked (or 'Beilby') films, and adsorbed films. All these influence the contact resistance and frictional characteristics of metal surfaces (Holm 1946, Bowden and Tabor 1950, Hirst and Lancaster 1954). In general the effect of these films is most marked when the load between the surfaces is small. Whitehead (1950) has extended the work of Bowden and Tabor (1945) to the study of metallic friction at very light loads. He found that in general Amontons' law, which states that the frictional force is independent of the area of the surfaces and directly proportional to the normal load, was obeyed over a very wide range of loads, but that significant variations in the coefficient of friction, which he attributed to the influence of the natural oxide layer, did occur. He suggested that this oxide layer might remain intact at light loads, and that the sliding under such conditions would be that of surfaces covered by oxide layers. The mechanism of sliding friction between two such oxide layers is similar to that between metal surfaces; oxide junctions can be formed and sheared, thus giving rise to the frictional force, but no metallic junctions occur until the applied load is sufficient to cause rupture of the natural oxide layers. These conclusions were based on measurements of the coefficient of friction and observations of the appearance of the track on a limited number of metals.

This paper describes experiments, carried out from 1950 to 1952, in which the measurement of contact resistance was introduced as a method of obtaining further information about the breakdown of the natural oxide. In addition, Whitehead's work was extended to cover most common metals. A brief note on the effect of oxide films at low loads has been published (Wilson 1952). The presence of an unbroken oxide layer greatly reduces surface damage, although it does not necessarily reduce the friction.

A comprehensive theory of the contact resistance between stationary surfaces was developed by Holm (1929) and his conclusions have been applied to sliding surfaces by Bowden and Tabor (1939). When an electric current passes between two clean metal surfaces the flow of charge is restricted to the areas of real contact, and this constriction appears as an extra resistance to the current. In the case of two flat surfaces there will be several areas of contact but if one of the surfaces is flat and the other spherical the contact is restricted to a single area. For a single clean contact area of radius a, Holm showed that $R=(\rho_1+\rho_2)/4a$ where ρ_1 and ρ_2 are the specific resistances of the metals. In practice most metal surfaces are covered by oxide layers and other contaminant films, which have a very high resistance. This high resistance part is proportional to $1/a^2$.

The surfaces near the metallic contact area may be partly conducting due to the tunnel effect. Calculations made by Holm show that even in the most favourable circumstances this cannot increase the current by more than about 2%. The contact resistance can thus be used to determine the area of metallic contact, and the value so calculated may be compared with the total area of contact, metallic and non-metallic, measured by direct microscopic examination.

This work has shown that measurement of the contact resistance provides a very sensitive method of detecting the breakdown of an oxide layer and of studying this effect quantitatively.

§ 2. EXPERIMENTAL

All the experiments were carried out on the light load friction apparatus designed by Whitehead which was modified in order to increase its scope and simplify the interpretation of results. The most important of these was an electrical device which enabled the contact resistance between the sliding surfaces to be measured and recorded simultaneously with the frictional force.

2.1. *The Friction Apparatus*

Figure 1 shows the essential parts of the apparatus. The block which rotates at a speed of about 0·01 cm sec^{-1} was clamped to a turntable driven by a synchronous motor. The slider which was screwed on to a spring steel suspension wire was a 45° cone with a rounded tip of radius 0·05 mm. The suspension wire was attached to a beam which was pivoted on a torsion bar and had its remote end supported on a micrometer. The lever arm ratio was unity and a predetermined normal load was applied to the slider by raising the remote end of the beam. The vertical restoring force in the suspension wire loaded the slider against the block. Seven interchangeable suspension wires enabled the load on the slider to be varied between 3 mg and 200 g. When the turntable carrying the block was set in motion, the slider moved with it in a horizontal plane until the restoring force was sufficient to shear the junction between the slider and the block. The lateral deflection of the wire, representing the frictional force, was measured by means of an electromagnetic pick-up.

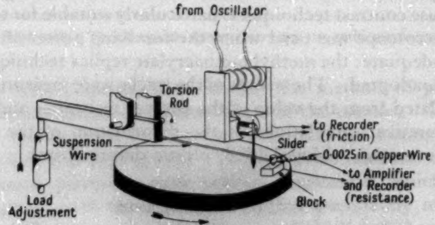


Figure 1. Sketch of essential parts of friction apparatus.

2.2. *The Resistance Apparatus*

The electrical resistance between the sliding surfaces was calculated from the measured value of the potential drop across the interface when an alternating 40 c/s current, drawn from a high impedance source, was passed between the block and slider. In order to make heating effects at the interface negligibly small, the current was adjusted so that the contact voltage did not exceed about 250μv, a value very much lower than the lowest value (70 mv for tin) recorded by Holm (1946) in his results of the measurements of the contact voltage required to produce thermal softening of the junctions which form the actual contact.

The contact voltage was amplified and measured with accuracy to ±3%; the apparatus was calibrated by inserting resistances of known value in place of the metal junction.

2.3. *Preparation of Surfaces*

It is almost impossible to prepare a series of standard surfaces which have identical characteristics. Recent developments in electropolishing techniques have made possible for the first time the preparation of relatively smooth stress-free surfaces on metals; however, even electropolished surfaces vary slightly from one to another. Electron micrographs indicate that electropolished surfaces, although much smoother than hand-polished ones, are still rough on an atomic scale. Fortunately surface roughness does not greatly influence the frictional characteristics of metals. Electropolishing techniques were used whenever possible to prepare both contacting surfaces. Except where otherwise stated the blocks were given a prolonged electropolish in order to remove all local surface stresses. The sliders, however, could be given only a light electropolish as the contour of the tip had to be preserved.

Since, as is shown in § 4.2, a worked surface is harder than a stress-free surface, the sliders were harder than the blocks, and scored clearly defined tracks.

When the surfaces had been suitably prepared they were subjected to a standard degreasing procedure consisting of vapour degreasing followed by electrolytic degreasing in 5% potassium hydroxide in methyl alcohol. The sliders, blocks, and their holders were considered grease free if they were completely wetted by water. They were tested before and after use.

2.4. *Examination of Surfaces*

The tracks produced by the sliders were examined using a metallurgical microscope; phase contrast technique is particularly suitable for this examination. The electron microscope was used when the resolving power of the light microscope proved inadequate: the methyl methacrylate replica technique developed by Brown (1950) was adopted. The widths of the tracks were measured and compared with those calculated from the value of the contact resistance and also with those deduced by theoretical calculations of the deformation of the contacts (Hertz 1881, elastic deformation; Meyer 1898, plastic deformation).

§ 3. THE OXIDE LAYER

Except when otherwise stated, the experiments described in this section were carried out on annealed specimens. Simultaneous records of the contact resistance and the frictional force were obtained for a large number of metals using loads varying from 3 mg to 200 g. The metals examined ranged from gold, which does not form an oxide in air, to magnesium and iron which form quite thick oxides under normal atmospheric conditions. The breakdown of the oxide was investigated by comparing the track width calculated from the contact resistance with that measured microscopically.

3.1. *Measurements of the Contact Resistance*

(*a*) *The Noble Metals.*

Gold does not form an oxide layer under normal conditions and thus measurements of its contact resistance provide an indication of the effect of adsorbed layers. It was found that the agreement between the track width calculated from the averaged contact resistance measurements and the optically measuerd values was quite good over the load range 0·2 g to 200 g (see figure 2). At very light loads the tracks could not be located, but the values calculated from contact resistance measurements did not fall on the theoretical curve. This trend was unchanged when the specimen was heated to 120°c and thus it is unlikely that adsorbed water molecules have any influence on the values obtained. Holm (1946) found that adsorbed water had a negligible influence on contact resistance. Went (1941) measured the resistance of static gold contacts and observed a similar increase in contact resistance. He calculated that a layer of adsorbed oxygen two molecules deep would be sufficient to account for this behaviour.

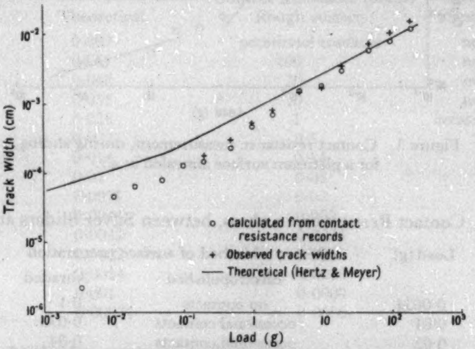


Figure 2. Track width measurements on gold surfaces.

Experiments with platinum showed that the method of preparation of its surface had a marked influence on the nature of the contact. This is in general agreement with the findings of other workers, see, for example, Holm (1946) and Güntherschulze and Betz (1938). Three types of surface were investigated. Annealed outgassed platinum gave contact resistances which except at the lightest loads agreed with theoretical values for completely metallic junctions (compare gold). Platinum which had been polished and then annealed in air gave the values shown in figure 3. The contact at loads above 1 g is essentially metallic since for annealed metals Meyer showed that $R \propto W^{-1/2 \cdot 6}$ where W is the load. With platinum at loads below 1 g $R \propto 1/W$; since $W \propto a^2$ this suggests that there is a comparatively thick layer of semiconducting material between the two surfaces. Finally, a specimen was outgassed *in vacuo* and then made an anode in sulphuric acid solution. The slope of the relation between contact resistance and load indicated that a semiconducting layer had again been produced. The contact resistance of all three surfaces remained finite at all loads used.

Silver surfaces were prepared in two ways : by electropolishing, which gave a very smooth finish, and by abrading on 600 grade carborundum paper under water, which gave a clean but rough surface. The contact resistances of these surfaces are given in table 1. It is apparent that the method of preparation of the surfaces can have a profound effect on the load at which breakdown of the oxide occurs.

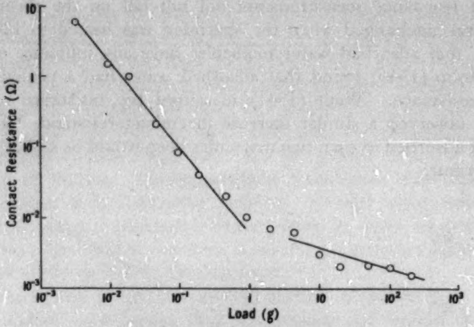


Figure 3. Contact resistance measurement, during sliding, for a platinum surface annealed in air.

Table 1. Contact Resistance, in ohms, between Silver Sliders and Blocks

Load (g)	Method of surface preparation	
	Electropolished	Abraded
0·0034	no contacts	0·1
0·01	occasional contacts	0·075
0·02	occasional contacts	0·04
0·05	5	0·03
0·1	5	0·01
0·2	0·5	0·015
0·5	0·3	0·005
1	0·1	0·0035
2	0·01	0·0025
5	0·0015	0·0015
10	0·001	0·001
20	0·0005	0·0005
50	0·0003	0·0003

Palladium behaves similarly to silver. The contact resistance measurements showed that it forms a definite oxide in air, and that the load at which this oxide breaks down depends on the surface finish. The presence of this oxide makes it difficult to decide whether an oxygen-containing layer, such as that detected on platinum, is also present in the surface layers of palladium.

(*b*) *The Base Metals.*

Contact resistance measurements were made during sliding between most common base metals. All these metals are known to form an oxide layer in air. It was demonstrated that, with the exception of indium, sliding can occur at light loads on all these metals without this oxide layer being ruptured. As the load is increased breakdown of the oxide begins, and the contact resistance begins to fluctuate violently. With many metals, such as iron and magnesium, this behaviour continued up to the heaviest loads used. With others, such as nickel, lead, and cadmium complete breakdown of the oxide occurs and the contact resistance trace becomes smooth; the resistance values then correspond to complete metallic contact. The actual load range over which breakdown occurs depends very markedly on the surface finish. Measurements on zinc, given in table 2, emphasize this point. On an abraded surface some breakdown of the oxide layer occurred at loads as low as 0·01 g and breakdown was practically complete at loads of 1 g, whereas, on an electropolished surface, breakdown began at 0·2 g and was not complete at 200 g. Consequently it is only possible to compare the loads at which oxide breakdown begins and is substantially complete for various metals if all the surfaces have been prepared by the same technique. Table 3 gives the results obtained for metals which had been prepared by electropolishing. The hardness of various metals and their oxides is also given. These values were obtained by micro-hardness measurements on natural or synthetic crystals of the oxides concerned, and showed a considerable scatter; they should be regarded as approximate only. The effect of the thickness of the oxide layer was also investigated.

Table 2. Influence of Surface Roughness on the Contact Resistance of Zinc

Load (g)	Contact Resistance (ohms)		
	Theoretical	Rough surface	Smooth surface
0·0034	0·087	occasional contacts	no contacts
0·01	0·061	100	no contacts
0·02	0·048	20	no contacts
0·05	0·035	10	no contacts
0·1	0·028	1	occasional contacts
0·2	0·022	0·5	100
0·5	0·014	0·14	30
1	0·01	0·05	2
2	0·0071	0·02	1
5	0·0044	0·007	0·07
10	0·0032	0·0035	0·10
20	0·0022	0·003	0·02
50	0·0014	0·0013	0·008
100	0·001	0·0009	0·007
200	0·00071	0·0006	0·0025

It was observed that if the motion of the slider was arrested during sliding on smooth electropolished surfaces in the load range where some oxide breakdown was occurring, the contact resistance fell to a very low value compared with that recorded during sliding. This behaviour suggests that as long as the slider keeps moving some cracking may occur although the oxide does not disintegrate completely, but that if the movement ceases the slider may penetrate the oxide due to creep deformation of the underlying metal.

(*c*) *Dissimilar Metals.*

All the results described so far refer to contact between two pieces of the same metal. A few representative combinations of dissimilar metals have also been studied. Whenever possible, combinations of metals having similar specific conductivities have been chosen, since this permits the direct comparison of contact resistance values.

Table 3. Measured Hardnesses of Various Metals and Oxides and the Load at which Oxide Breakdown begins on Electropolished Surfaces

Metal	Hardness (kg mm^{-2})	Hardness of oxide (kg mm^{-2})	Load at which breakdown begins (g)
gold	20	—	—
platinum	40	—	—
silver	26	—	0·003
palladium	100	—	0·02
indium	1	—	—
tin	5	SnO_2 1650 SnO 350	0·02
lead	5	250	0·1
cadmium	20	—	0·1
aluminium	15	1800	0·2
molybdenum	220	50	0·2
cobalt	240	—	0·2
zinc	35	200	0·5
copper	40	130	1·0
nickel	86	430	1·0
magnesium	27	MgO 400 $Mg(OH)_2$ 150 $MgCO_3$ 50	10·0
iron	105	all about 400	10·0
chromium (plated)	850	1600	above 500

Copper and gold. This demonstrates the nature of the contact between two metals with similar physical properties, one of which does not form an oxide in air. The hardness of the annealed gold specimen was 24 kg mm^{-2} and that of the annealed copper 38 kg mm^{-2}. The contact resistance values are given in table 4. The values for gold on gold can be regarded as representative of those obtained between oxide free surfaces, but in comparing these results it must be remembered that since the specific resistance of gold is 50% higher than that of copper a clean contact on gold has a higher ohmic resistance than one of the same area on copper. When copper slides on gold the oxide on the copper

Table 4. Contact Resistance between Gold and Copper Sliders and Blocks

Load (g)	Contact Resistance (ohms)			
	Gold on gold	Copper on gold	Copper on copper	Gold on copper
0·0034	1·0	no contact	no contact	no contact
0·01	0·05	no contact	no contact	no contact
0·02	0·035	no contact	contact at start	no contact
0·05	0·03	occasional contact	10	occasional contact
0·1	0·0175	1·0	1	20
0·2	0·015	0·75	2	4
0·5	0·01	0·005	0·5	1
1	0·005	0·002	0·1	0·1
2	0·0035	0·0015	0·02	0·01
5	0·0015	0·001	0·005	0·007
10	0·0013	0·00085	0·001	0·0035
20	0·0008	—	0·00075	0·0015
50	0·0004	—	0·0005	0·0013
100	0·0003	0·00035	0·0004	0·00125
200	0·0002	0·0003	0·0003	0·0014

slider prevents metallic contact at light loads. As the load increases the oxide breaks down, and finally the resistance values become very close to those corresponding to contact between clean surfaces. When gold slides on copper the gold slider, although oxide-free itself, always encounters unbroken copper oxide, and, as it is softer than the copper, it does not disrupt the oxide layer as much as a copper slider on copper.

Platinum and tin. This illustrates the contact between two metals of different hardness one of which does not form an oxide in air. The platinum used in these experiments was about ten times harder than the tin. Tin forms a hard oxide which is readily disrupted. Contact resistance measurements showed that there was considerable contact when platinum slid on tin. The hard platinum slider penetrates the thin brittle oxide layer on the soft tin very readily. The contact resistance values were of the same order as those for tin on tin. The contact resistance values for tin sliding across a platinum surface were much higher than those recorded for tin or for platinum on tin.

Lead and steel. This pair is a typical example of two metals one of which is much softer than the other and both of which form an oxide layer.

The steel had a hardness of 140 kg mm^{-2}, and the lead a hardness of 5 kg mm^{-2}. The contact resistance values for the various combinations are given in table 5 :

Table 5. Contact Resistance between Lead and Steel Slides and Blocks

Load (g)	Resistance (ohms)			
	Lead on lead	Lead on steel	Steel on steel	Steel on lead
0·0034	10	no contact	no contact	no contact
0·01	2	no contact	no contact	no contact
0·02	0·5	no contact	no contact	no contact
0·05	0·2	no contact	1000	50
0·1	0·1	no contact	70	8
0·2	0·2	no contact	8	1
0·5	0·05	100	7	0·7
1	0·025	50	8	0·5
2	0·023	40	8	0·3
5	0·022	8	8	0·1
10	0·018	8	7	0·05
20	0·016	3	2	0·04
50	0·015	0·5	0·1	0·03
100	0·015	0·03	0·075	0·025
200	0·012	0·01	0·05	0·02

the contact resistance values for steel sliding on lead are lower than for steel on steel; on the other hand the values for lead on steel are, in general, higher than those for lead on lead. In no case was the contact completely metallic.

3.2. *Measurements of the Coefficient of Friction*

The value of the coefficient of friction and the nature of the sliding between oxide layers would be expected to differ from that observed for metallic surfaces. However, it was found that the appearance of the track does not necessarily change when the oxide breaks down, although it may show marked changes in appearance at loads not related to those required to disrupt the oxide. Such a change in the nature of the sliding, which was also associated with a marked change in the coefficient of friction, was observed on annealed gold, which has no oxide layer.

Experiments showed that only fully work-hardened metals obey Amontons' law. Comparisons of the values for the coefficient of friction of metal and oxide have been made on such surfaces. Figure 4 shows that the coefficient of friction for oxide sliding on oxide can exceed, equal or be less than that between the corresponding metal surfaces. Sliding between oxide layers is always relatively smooth, but when breakdown of the oxide occurs the motion may become irregular.

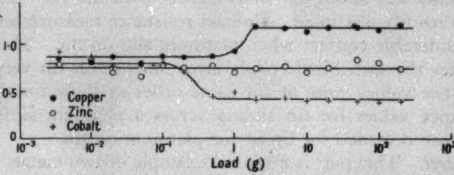


Figure 4. Coefficient of friction measured on fully work-hardened surfaces. Oxide breakdown begins at *c.* 1 g load in copper, *c.* 0·5 g in zinc and *c.* 0·2 g in cobalt.

In no case was the air-formed oxide layer thick enough to see, even when the electron microscope was used, but an idea of the physical mechanism of the penetration of an oxide layer was obtained by observing the disruption of an anodized layer of aluminium. This was produced by a sulphuric acid anodizing process, and had an amorphous structure resembling that of the natural air-formed layer. It was, however, at least one thousand times thicker than the natural layer. Figure 5 (Plate I) shows the stages in the breakdown of this anodized layer. At light loads the oxide remains intact; as the load increases a regular crack pattern is produced, with the cracks at an angle of about 45° to the direction of sliding. As the load increases further cracks develop across the base of the tracks and, finally, the oxide layer is penetrated.

3.3. *The Effect of the Oxide Layer*

The oxide layer which is present on most metals under normal conditions was found to have a profound effect on the nature of the contact between their surfaces. At light loads the oxides slide over each other without being disrupted; the properties of the junction are then controlled by the oxides and the electrical resistance of the contact is very high. As the load is increased the deformation of the contacting surfaces ruptures the oxides and metal-to-metal contact occurs; the properties of the junction become determined by those of the metal and the contact resistance falls to a very low value.

The appearance of the track is not a reliable guide as to whether the sliding occurred between oxide and oxide or metal and metal. Sliding between oxides usually produces a smoother track than that between metals, but this is not always so; with lead and nickel, for example, there is no visible change in the track when the oxide is penetrated (see figure 6 (Plate II)).

Contact resistance measurements on gold showed that, in air, a very thin layer of impurity, probably oxygen, is adsorbed. This adsorbed layer affects the contact resistance only at the very lightest loads and there is no evidence that gold forms an oxide in the normal sense.

Platinum which had been outgassed *in vacuo* at a high temperature behaved similarly to gold; but when more usual methods were used to prepare the surfaces the contact resistance was found to be greater than that calculated for completely metallic contact, and to depend on the load in a manner indicating the presence of a semiconducting surface layer. There are two ways in which such a layer might be produced. It could be a residual Beilby layer; when a polished surface, on which there is an amorphous layer, is annealed all traces of the layer do not disappear. Moore and Tegart (1952) suggested that the residual effect is due to the presence of extraneous particles which have been embedded in the surface during the polishing operations. These particles may interfere with the passage of electrons and so increase the specific resistance of the surface layers. The main source of these extraneous particles is thought to be oxide which is included into the surface as a result of the deformation; with platinum, however, there may be no oxide available.

An alternative explanation is that the surface layers contain an appreciable amount of adsorbed oxygen (Sidgwick 1950). The oxygen atom is large and might be expected to interfere with the electrical conductivity.

Measurements on platinum after it had been made the anode of an electrolytic cell give strong support to the view that oxygen can be absorbed into the surface layers of platinum and produce a semiconducting layer. The contact resistance values recorded when this layer is present, although higher than those obtained on a clean metal surface, are nevertheless much lower than would be expected if a distinct oxide layer was formed.

All other metals examined formed a distinct oxide layer in air which had a very much higher conductivity than the parent metal.

In contact between similar metals the load at which breakdown of the oxide begins depends largely on the relative mechanical properties of the oxide and the underlying metal. Thus, if a thin hard brittle oxide covers a soft metal it will be almost completely disrupted by very light loads: on the other hand, if the metal and its oxide have similar mechanical properties the oxide will deform in a similar manner to the metal and will impede metallic contact even at great loads. The behaviour of tin is typical of the first case, and that of iron and, to a less extent, copper of the second.

The load at which breakdown occurs is influenced also by the surface roughness and the thickness of the oxide. With rough surfaces breakdown occurs more readily than with smooth ones; on the other hand, metallic contact can be reduced to some extent by deliberately thickening the oxide. The results do not indicate any dependence of the breakdown on the crystallographic orientation of the underlying metal.

When the difference in hardness of the contacting metals is great another effect becomes important. A hard slider moving over a soft block may rapidly become coated with the soft metal and sliding will then be between two surfaces of the soft metal. A substantial proportion of the oxide at the original interface may remain intact and make a permanent contribution to the contact resistance. The experiments with a steel slider moving over a lead block illustrate this effect.

These results support the picture of the formation and growth of junctions put forward by MacFarlane and Tabor (1950).

§ 4. THE BEILBY LAYER

4.1. *Measurement of the Contact Resistance*

It is now generally accepted that there is a layer on a mechanically polished metal surface very similar to that originally envisaged by Beilby (1921). There is very little information concerning the effect of this layer on contact resistance. Chambers and Pippard (1953) measured the resistance of metal surfaces using high frequencies and concluded that the Beilby layer has only a slight influence on the electrical conductivity.

No effects which could definitely be attributed to the presence of the Beilby layer were recorded.

4.2. *Measurement of the Coefficient of Friction*

The Beilby layer and the worked metal associated with it can have a very marked effect on friction in the range of loads under consideration. Finch, Quarrell and Roebuck (1934) showed that the surface layers of a mechanically polished surface are amorphous, and Burwell and Wulff (1939) showed that, although this amorphous layer is only a few hundred ångströms thick, the deformed layer beneath can extend for many thousands of ångströms. There is, of course, no clear demarcation between these two layers.

On some metals changes in the coefficient of friction and in the appearance of the track were detected which were not associated with breakdown of the oxide layer. These changes were confirmed by experiments on gold and outgassed platinum surfaces which have no oxide. The results of frictional measurements on the annealed gold and platinum are given in figure 7. There is a change in the appearance of the track in the load range where the coefficient of friction changes. At light loads the sliding is regular and, although the coefficient of friction is fairly high, the track is smooth. The coefficient of friction rises at greater loads, and the motion becomes erratic and heavy scoring occurs.

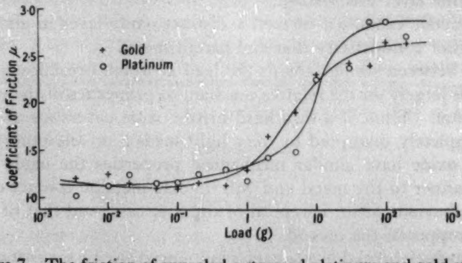


Figure 7. The friction of annealed outgassed platinum and gold surfaces.

Experiments were also carried out on gold and platinum surfaces which had been prepared by other methods. In one experiment, an annealed gold surface was abraded on 600 carborundum and then polished by hand so that it acquired a Beilby layer. This treatment reduced the coefficient of friction below that of a fully annealed surface. Platinum behaved similarly. Finally a fully work-hardened platinum specimen was examined. It was found that in this case Amontons' law was obeyed and that the coefficient of friction was lower than that recorded with platinum specimens prepared by other methods.

Micro-hardness tests provide a simple non-destructive means of measuring the extent of the work-hardened layers. Instruments are now available which, when used with great care, will make measurements at loads down to 1 g. It is possible to make impressions less than one micron deep and, by regularly increasing the load, the degree and extent of the surface hardening may be determined. Results obtained in this way on the platinum surfaces used in the friction experiments are plotted in figure 8. The increase in surface hardness when an annealed specimen is hand-polished is quite pronounced. The results on gold were similar.

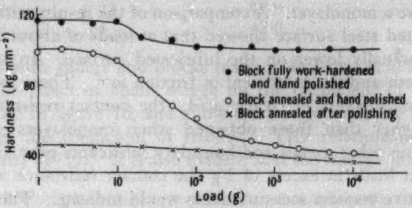


Figure 8. Micro-hardness measurements on platinum surfaces.

With the base metals the exact behaviour is obscured by the presence of an oxide layer, although some changes in the coefficient of friction not directly connected with oxide breakdown are observed with, for example, lead and nickel (figure 6).

4.3. *The Effect of the Beilby Layer*

It has been demonstrated that for many metals Amontons' law is not obeyed. Contact resistance records show that this is not entirely due to the breakdown of their oxides as suggested by Whitehead. In fact the nature of contact is markedly dependent on the load even with surfaces on which no oxide can be detected. The effect of the Beilby layer, although usually masked by that of the oxide layer, may be important on surfaces which have no oxide.

If an annealed mechanically polished metal touches another metal lightly, the nature of the contact will be determined by the properties of the surface layer and will be characteristic of contacts between work-hardened metals. At heavier loads however the properties of the bulk metal become more important and the contact becomes typical of those between annealed metals. Measurements on hand-polished annealed gold and on outgassed platinum illustrate this.

§ 5. ADSORBED LAYERS

Boundary lubricants are probably the most important class of adsorbed layers. It is known that to perform effectively boundary lubricants must form solid layers on a metal surface. Nevertheless it has been shown by electrographic methods (Moore 1948) and by radioactive transfer techniques (Rabinowicz and Tabor 1951) that some metallic contacts are always formed even through the best boundary lubricants, and consequently an attempt was made to make quantitative measurements using the contact resistance method.

5.1. *Measurements of the Contact Resistance*

The contact occurring during sliding of two platinum surfaces lubricated by a stearic acid monolayer applied by retraction from solution was studied by means of the contact resistance method. The platinum behaved very much as though it was covered by an oxide layer; at light loads (below 100 mg) there was no contact during sliding, but as the load increased there was some intermittent metallic contact and, finally, at loads above 50 g the contact resistance became comparable with that recorded between clean surfaces.

A similar experiment was conducted on steel. In this case the stearic acid was applied from the melt and it probably reacted to form a stearic soap before retracting to leave a monolayer. A comparison of the results with those obtained on an unlubricated steel surface showed that at loads of about 20 g the contact resistance was actually lower on the lubricated surface. In spite of this the sliding was smooth and the coefficient of friction low. These experiments were repeated using multilayers of stearic acid; the contact resistance values were only slightly higher than those obtained when monolayers were employed. Experiments using other metals and boundary lubricants gave similar results.

In general, at loads in excess of 1 g the contact resistance was much lower than the radioactive transfer measurements would indicate. This shows that the area of contact as defined by the area through which an electrical charge may flow and calculated from measurements of the constriction resistance is greater than the area of actual metal-to-metal contact as calculated from the observed frictional force or indicated by radioactive transfer techniques.

5.2. *Measurements of the Coefficient of Friction*

The frictional records showed that, regardless of nature of the underlying metal, the coefficient of friction was high when there was no electrical contact, and low when the contact resistance was very low, suggesting that this behaviour was characteristic of the lubricant used. In order to check this several experiments were carried out using sliders and blocks cut from solid lubricants.

Observations of the sliding between two stearic acid surfaces confirmed that the coefficient of friction was dependent on load. Some typical results are shown in figure 9. It will be seen that there is no tendency for the coefficient of friction to rise continually as the load is reduced as Whitehead suggested. Figure 9 also shows the values obtained when stearic acid slid over clean platinum, and over platinum covered by a retracted monolayer of stearic acid. It is apparent that the results shown follow the same general pattern. The shear strength of

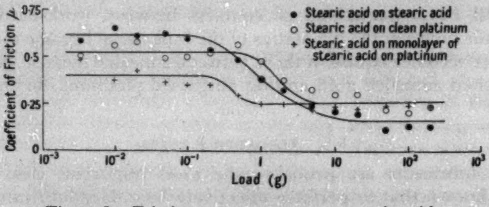


Figure 9. Friction measurements on stearic acid.

stearic acid was calculated by measuring the tracks left on the platinum surface by the stearic acid slider and assuming circular contact. Some typical values are given in table 6.

Table 6. Shear Strength of Stearic Acid

Load (g)	Track width (mm)	Friction force (g)	Shear strength (g mm⁻¹)
5	0·1	1·6	204
10	0·12	3·1	274
20	0·15	6·2	351
50	0·2	10·5	334
100	0·27	17·0	297
200	0·4	42·0	335

Experiments with sliders and blocks made from lauric acid and octacosanoic acid illustrate the effect of chain length on the coefficient of friction. These results are plotted in figure 10, and show that, in general, the longer the chain length the lower the coefficient of friction. This relationship was originally observed by Hardy (1936). It is also evident that the change from a higher value to a lower value of the coefficient of friction begins at lighter loads with the longer chain acids. As with stearic acid, the track widths were measured and used to calculate the shear strengths of the acids. The values were all between 200 and 300 g mm^{-2}, which is in good agreement with those obtained by Bailey and Courtney-Pratt (1955).

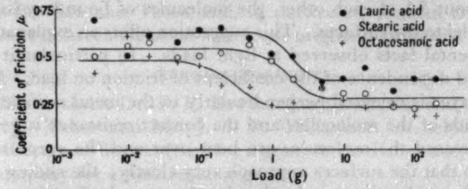


Figure 10. Effect of chain length on the friction of fatty acids.

Sebacic acid is a dibasic acid containing the same number of carbon atoms as stearic acid. As it has an active group at each end of the molecule it takes up an orientation parallel to a metal surface. In this respect it differs from stearic acid which orients itself almost perpendicular to a metal surface. Measurements of the frictional force were made for sebacic acid sliding on itself and also on a platinum surface. In both cases the coefficient of friction remained low and constant at all loads.

A number of experiments on various metal soaps suggested that they behave similarly to the fatty acids. A few straight chain hydrocarbons were also examined. Unfortunately these had a tendency to crumble in the heavier load ranges, but the results followed the same general pattern as with fatty acids.

5.3. *The Effect of Lubricant Films*

The contact resistance measurements have shown that even the most effective boundary lubricants are able to resist the close approach of metal surfaces only at loads below one gramme; at greater loads very low contact resistance values, often of the same order as those obtained on unlubricated surfaces, were recorded. On the other hand the measurements of the frictional force, and also the radioactive transfer investigations referred to earlier, indicate that with an effective boundary lubricant there is very little true metallic contact. Consequently the area conducting electricity must be greater than the area over which the metals actually touch. It is suggested that at loads above one gramme the metal surfaces must be very close together; so close in fact that most of the conduction may be explained by the tunnel effect. Calculations show that under the experimental conditions employed the conduction would be significant only if the surfaces approached to about 5 Å over all the area which undergoes plastic deformation, that is over all the region which would, but for the presence of the lubricant, be in true contact. At the same time of course, the surfaces would have to be free from any insulating oxide layer. Two possible mechanisms by which the oxide layer could be removed are mechanical abrasion during sliding, and chemical attack by the more active boundary lubricants. In either case the metal would remain clean only because it would be protected from renewed oxidation by the presence of the boundary lubricant itself.

If, as the contact resistance measurements indicate, the metal surfaces do approach to about 5 Å of each other, the molecules of boundary lubricant must be almost parallel to the surfaces. This suggestion offers an explanation for many of the experimental facts observed at light loads. In particular it can account for the observed dependence of the coefficient of friction on load. At light loads the boundary layer is oriented perpendicularly to the metal surface; the sliding is across the ends of the molecules, and the contact resistance is very high. As the load is increased the molecules are bent over and the surplus lubricant is squeezed out so that the surfaces approach very closely; the sliding is now along the length of the molecules, conduction by the tunnel effect is possible. The transition from one type of sliding to the other occurs at a lower load when a long chain fatty acid is used than with a short chain length acid. With a dibasic acid the friction is low and constant at all loads since the acid is always oriented parallel to the surface.

Table 7. Change in Mean Pressure on Stearic Acid Tracks with Increase in Load

Load (g)	Track width (mm)	Pressure (g mm^{-2})	Coefficient of friction
5	0·1	640	0·32
10	0·12	880	0·31
20	0·15	1130	0·31
50	0·2	1600	0·21
100	0·27	1720	0·17
200	0·4	1600	0·21

It is not at first evident why increasing the load should in itself cause a bending over of the molecules; if the mean pressure were constant this would not be expected. However, calculations made from measurements of the tracks left by fatty acid sliding on platinum show that the mean pressure on the acid increases as the load increases, see table 7. Thus experiments indicate that the pressure increases with load and that when it exceeds about 1000 g mm^{-2} there is a fall in friction. It is suggested that this corresponds to the bending of the molecules.

§ 6. Conclusions

The behaviour of metals when in sliding contact has been studied by making simultaneous measurements of contact resistance and friction under a wide range of conditions and the effect of various surface films has been investigated. When one metal slides over another the frictional force is the force necessary to shear the junctions formed between the two surfaces. In the simplest case, that of a clean, oxide-free, fully work-hardened metal, the junctions have similar mechanical properties to the bulk. If the mechanical properties of the surface layers of the metal differ from those of the bulk metal, or if the surface layers acquire different properties during the formation of the junctions, the size of the areas of contact depends on the properties of both the bulk metal and the surface layers. This dependence obeys certain specific laws and the behaviour can be explained by a simple extension to the adhesion theory of friction. If oxide or lubricant films are present the growth of metallic junctions is restricted, and at light loads may be completely prevented. In this latter case the frictional force is the force necessary to shear the junctions formed between the oxide or lubricant films.

The results are in accord with the general theory of contact between surfaces developed by Bowden and Tabor and, in particular, confirm and amplify the picture of contact between metals at light loads originally proposed by Whitehead.

Acknowledgments

I should like to thank Dr. F. P. Bowden for his interest and encouragement and also Dr. D. Tabor and Mr. J. B. P. Williamson for helpful discussions. I thank the Ministry of Supply (Air) for a grant to the Laboratory, and the Shell Petroleum Company for a maintenance grant.

References

Bailey, A. I., and Courtney-Pratt, J. S., 1955, *Proc. Roy. Soc.* A, **227,** 500.
Beilby, G., 1921, *Aggregation and Flow of Solids* (London : Macmillan).
Bowden, F. P., and Tabor, D., 1939, *Proc. Roy. Soc.* A, **169,** 391; 1945, *Ann. Reports Chem. Soc.*, **42,** 20; 1950, *The Friction and Lubrication of Solids* (Oxford : University Press) (Revised version 1954).
Brown, A. F., 1950, *Monogr. Ser. Inst. Metals*, No. 8, p. 157.
Burwell, J. T., and Wulff, J., 1939, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **135,** 486.
Chambers, R. G., and Pippard, A. B., 1953, *Monogr. Ser. Inst. Metals*, No. 13, p. 281.
Finch, G. I., Quarrell, A. G., and Roebuck, J. S., 1934, *Proc. Roy. Soc.* A, **145,** 676.
Güntherschulze, A., and Betz, H., 1938, *Z. Electrochem.*, **44,** 253.
Hardy, W. B., 1936, *Collected Works* (Cambridge : University Press).
Hertz, H., 1881, *J. reine angew. Math.*, **92,** 156.
Hirst, W., and Lancaster, J. K., 1954, *Proc. Roy. Soc.* A, **223,** 324.
Holm, R., 1929, *Wiss. Veröff. Siemens*, **7,** 217; 1946, *Electric Contacts* (Uppsala : Almqvist and Wiksells).
MacFarlane, J. S., and Tabor, D., 1950, *Proc. Roy. Soc.* A, **202,** 224.
Meyer, A., 1898, *Öfvers. VetenskAkad. Förh., Stockh.*, **55,** 199.
Moore, A. J. W., 1948, *Proc. Roy. Soc.* A, **195,** 231.
Moore, A. J. W., and Tegart, W. J. McG., 1952, *Proc. Roy. Soc.* A, **212,** 458.
Rabinowicz, E., and Tabor, D., 1951, *Proc. Roy. Soc.* A, **208,** 455.
Sidgwick, N. V., 1950, *The Chemical Elements and Their Compounds* (Oxford : University Press), p. 1578.
Went, J. J., 1941, *Physica*, **8,** 233.
Whitehead, J. R., 1950, *Proc. Roy. Soc.* A, **201,** 109.
Wilson, R., 1952, *Proc. Roy. Soc.* A, **212,** 450.

The Direct Separation of the Reversible and Irreversible Components of the Magnetothermal Effect

By L. F. BATES and N. P. R. SHERRY†
Department of Physics, University of Nottingham

MS. received 24th March 1955

Abstract. A new method is described for separating reversible and irreversible components of the magnetothermal effect by carrying out small, 'backward' increments in magnetization at given points on the hysteresis cycle. With the sensitivity of the existing apparatus this method is of limited application, but with increased sensitivity it could be applied generally. Results for the reversible and irreversible heat changes measured with a cobalt specimen, confirm that if reversible thermal effects alone are measured, the b'' coefficient, as defined in the Stoner and Rhodes theory, remains constant over the hysteresis cycle.

§ 1. INTRODUCTION

When a ferromagnetic material is taken around a hysteresis cycle there is always an energy dissipation, and in recent years measurements have been made of the several heating and cooling effects which occur in different parts of the cycle (see Bates 1951), in the hope of obtaining information about the energies associated with the several magnetization processes. As with most technical magnetization phenomena, an initial classification of magnetothermal effects simply consists of dividing them into reversible and irreversible effects. It is clear that such a division is only a beginning since further subdivision may be made when the particular energy changes associated with any one mechanism are understood. However, practical magnetothermal measurements of the reversible and irreversible quantities must first be made, but, as many writers have emphasized, the separation is important if full use is to be made of the measurements in an analysis of the various magnetization processes.

However, at present the state of the theory is such that irreversible effects can only be handled in a qualitative manner, and numerical predictions are therefore referred to a highly idealized material and are not applicable to practical cases except in general terms. Fortunately, the same is not true of reversible magnetothermal effects, and the Stoner and Rhodes (1949) theory is applicable to real materials. Results are presented below of what we believe are the first direct measurements of reversible heat changes, together with a discussion of how the measurements support the above theory.

§ 2. REVERSIBLE AND IRREVERSIBLE HEAT CHANGES

2.1. *Irreversible Changes*

An irreversible contribution usually occurs when the material 'jumps' from one metastable state to another. As the material is taken slowly around a hysteresis cycle, such contributions occur mainly as a coercive point, where large irreversible magnetization changes take place, is being traversed. A theory of Rhodes (1948), for an idealized ferromagnetic material, predicts that little irreversible heating will take place outside the steep part of the hysteresis cycle and that at a coercive point the heating rises sharply to the value of the total overall heating per half-cycle. In practice the heating step will not be as abrupt as this simple theory would suggest and a rounding off of the corners of the curves of thermal changes occurs. It is generally thought that irreversible thermal curves can manifest themselves as a heating of the specimen only, but it has been recently suggested by Tebble, Wood and Florentin (1952) that in certain cases an irreversible cooling may be possible. Perhaps one might here draw attention to the extraordinary magnetothermal behaviour of $MnAu_2$, described by Meyer and Tagland (1954).

---

† Now at Services Electronics Research Laboratory, Harlow, Essex.

2.2. *Reversible Changes*

When reversible thermal effects are considered, the quantitative thermodynamical treatment of Stoner and Rhodes can be applied. Briefly, they divide a heat change $\Delta Q'$ into two parts, which may be written as follows:

$$Q' = a\int d(IH) + b\int H dI. \qquad \ldots\ldots(1)$$

The two coefficients a and b are dependent on the material under investigation and are given by

$$a = -\frac{I}{I_0}\cdot\frac{dI_0}{dT} \quad \text{and} \quad b = \frac{I}{K}\cdot\frac{dK}{dT} \qquad \ldots\ldots(2)$$

where T is the temperature, I_0 the intrinsic magnetization and K the principal anisotropy coefficient. The symbols I and H refer respectively to the intensity of magnetization of the specimen and the effective field to which it is subjected. The first term on the right-hand side of equation (1) is due to changes in intrinsic magnetization and reduces at high fields to the familiar expression for the magnetocaloric effect. The second term is the contribution to the energy effects of reversible rotational processes. Although other terms might occur in the expression, the above represent the chief energy processes, and other effects may be considered as second-order effects, e.g. the anomalous low-field effect (Bates and Sherry 1955).

§ 3. PREVIOUS SEPARATION OF REVERSIBLE AND IRREVERSIBLE HEAT CHANGES

It is clear that reversible and irreversible heat changes must be measured separately if the results are to be considered in the light of the above theory. When earlier 'total' thermal curves are examined, the Stoner and Rhodes theory is satisfactory in the regions of moderate and high fields, but it breaks down at low fields, because the measurements in the latter region include irreversible effects which are not considered in the theory.

There have been two attempts to measure reversible heat changes experimentally. The first was a direct method applied by Okamura (1936) to a large number of materials. But his method does not appear to be sound and one can only say that his results give an indication of the types of the two separate effects to be expected.

More recently an indirect measurement of heat changes has been developed at Leeds (Tebble, Wood and Florentin 1952). This is based on the simple thermodynamical relation

$$\Delta Q'_{\mathrm{rev}} = (\partial Q'/\partial H)_S \, dH = -T(dI/dT)_H \, dH \qquad \ldots\ldots(3)$$

and involves the measurement of small changes of the intensity of magnetization with temperature. So far, these are the only truly accurate measurements of reversible thermal effects, since any direct method must be based on approximations. However, these indirect readings must be supplemented by a direct measurement of the total thermal changes which cannot be calculated from equation (3). Since the latter measurement involves disturbing the specimen between the two sets of measurements, it appeared that it would be worth while to investigate the possibility of making a valid direct measurement of heat changes in such a way that the total and reversible curves could be obtained at the same time on the same specimen.

§ 4. PRESENT METHOD OF SEPARATION

The obvious way of measuring the reversible component of the thermal change is by a method which is analogous to that used in the measurement of reversible susceptibility, viz. by carrying out small backward and forward increments at given points on the hysteresis cycle. There seems little doubt that at moderate and high fields this operation may be satisfactory, but it has been shown that to apply this method in the low field or coercive point regions, extremely small increments in field would be necessary. The work leading to these conclusions was, however, carried out on materials of low coercivity with a hysteresis curve falling sharply in these regions. Thus, it is possible that in materials of high coercivity a satisfactory approximate separation might be effected by taking small backward increments on the cycle.

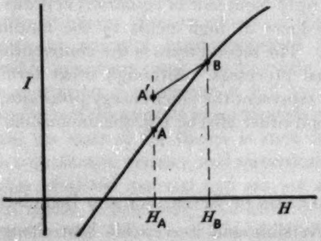


Figure 1. Method of separating reversible and irreversible changes.

The method employed may be followed by reference to figure 1 in which the unbroken curve depicts a portion of the magnetic hysteresis curve for a ferromagnetic material that is in the cyclic state and has arrived at the conditions represented by point A. When the field increment H_A to H_B is made, the thermal effect accompanying the change is measured by the Bates method which is fully described elsewhere (Bates 1951). This, of course, is the normal method of measuring the total thermal change for this step. Now, instead of continuing around the cycle, the field is returned to H_A and, unless the step AB is entirely reversible, the material will now be at a condition A′ which is different from A. If this subsidiary step is reversible, when the field is again increased to H_B a return should be made to B. Strictly, this is only true if the step AB is vanishingly small, but in practice finite changes of sufficient magnitude have to be taken to give a measurable thermal effect. The reversibility test to be applied here is that the heat change accompanying the step BA′ must be equal in magnitude but opposite in sign to that of A′B. A further check of reversibility consists in observing that the associated increments in magnetization are also equal and opposite. The field increments are so chosen that the difference between the magnitude of the thermal changes BA′ and A′B is smaller than the experimental error and therefore to a first approximation the area of the closed subsidiary loop is so small that irreversible effects can be neglected.

It is reasonable to assume that the step A′B represents the reversible part of the step AB since, as long as the subsidiary step is reversible, the measured $\Delta Q'_{\text{rev}}/\Delta H$ value cannot depart much from the true limiting value, and there are few irreversible jumps or the subsidiary loop would not be closed. Thus it appears that a satisfactory direct separation may be made when materials with a moderate coercive field are chosen.

§ 5. RESULTS AND DISCUSSION

The first measurements made on the above lines were made on a specimen of electrolytic cobalt. The sensitivity of the system was approximately the same as that of the Bates and Marshall system, viz. 500 erg cm⁻³ mm⁻¹ deflection. Reversible and total readings were made on specimens both in the heat-treated and untreated states, and the irreversible heat changes were deduced from the simple relation

$$Q'_{\text{tot}} = Q'_{\text{rev}} + Q'_{\text{irrev}}. \qquad \ldots\ldots(4)$$

The results are shown graphically in figures 2 and 3 respectively. Considering first the heat-treated specimen, figure 2, it is seen that the total curve Q'_{tot} exhibits

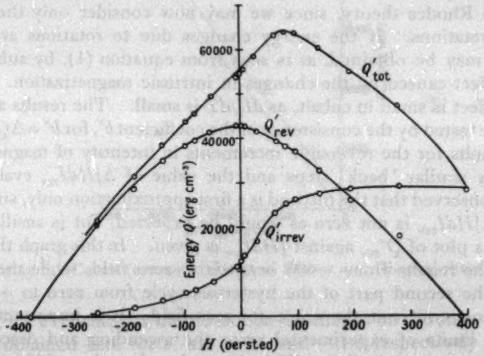


Figure 2. Cobalt (heat-treated)—separation of reversible and irreversible changes.

the usual form for cobalt (Bates and Harrison 1948): it rises slowly to a maximum and then descends to the value of the energy loss per half-cycle. The extreme portions of the graph are linear, because the function $\int HdI$ is linear over these regions. The overall heating is in good agreement with the value predicted by Warburg's law. The curve for irreversible changes Q'_{irrev} also appears to be satisfactory: because little change takes place at the larger field values, the major part of the heating occurs on the steep part of the hysteresis cycle, and the middle of the rise is at the coercive point (34 oersteds). No irreversible cooling is observed. The curve for reversible changes is symmetrical about the zero field axis, and the net thermal effect is only -1400 erg cm^{-3}. The latter may appear to cast doubt on the validity of the separation until it is remembered that this value is obtained from a summation of the $\Delta Q'_{\mathrm{rev}}$ values, and such a discrepancy is within the limits of experimental error in view of the observed maximum reversible heating of 42 800 erg cm^{-3}.

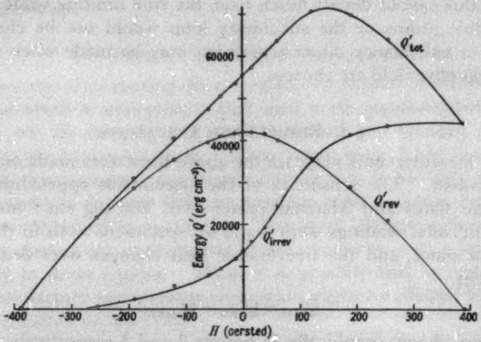


Figure 3. Cobalt (untreated)—separation of reversible and irreversible changes.

A quantitative test of these results can be obtained by the application of the Stoner and Rhodes theory, since we may now consider only the effect of the reversible rotations. If the energy changes due to rotations are denoted by Q''_{rev}, they may be obtained, as is seen from equation (1), by subtracting from Q'_{rev} the effect caused by the changes in intrinsic magnetization. Fortunately, the latter effect is small in cobalt, as dI_0/dT is small. The results and the theory may then be tested by the consistency of the coefficient b'', for $b'' = \Delta Q''_{\mathrm{rev}}/\Delta \int HdI_{\mathrm{rev}}$.

The results for the reversible increments in intensity of magnetization were obtained by similar 'back' steps and the value of $\Delta \int HdI_{\mathrm{rev}}$ evaluated. Once again, it is observed that the method is a first approximation only, since the overall value for $\Delta \int HdI_{\mathrm{rev}}$ is not zero as would be expected, but is small and positive. In figure 4 a plot of Q''_{rev} against $\int HdI_{\mathrm{rev}}$ is given. In this graph the open circles represent the results from -400 oersteds to zero field, while the black circles represent the second part of the hysteresis cycle from zero to $+400$ oersteds; the outermost end is at the value for zero field. It is important to note that within the limits of experimental error the ascending and descending points fall on the same straight line. This shows that the value of b'' is constant right down to the lowest fields. Previous results have indicated b'' to be constant at

the higher field values, but it is now shown for the first time that direct experimental support is obtained for the Stoner and Rhodes theory over the whole field range. The experimental results give $b'' = -1{\cdot}1$ which is in good agreement with the calculated value, particularly in view of the possible unreliability of the values of dK/dT.

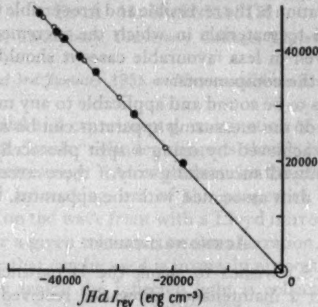


Figure 4. Cobalt (heat-treated)—test of Stoner and Rhodes' theory.

In the case of untreated cobalt, figure 3, the results again appear reasonable, and the b'' plot in figure 5 shows the coefficient b'' to be constant over a large part of the field range, but with a slight increase as zero field is approached. The actual value of b'' is again most satisfactory, as both the linear portions have a slope of $-1{\cdot}3$.

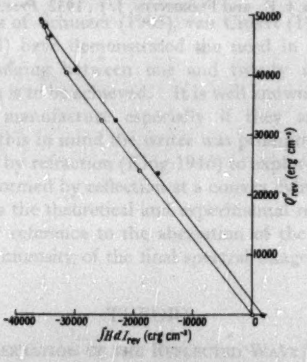


Figure 5. Cobalt (untreated)—test of Stoner and Rhodes' theory.

In cobalt, the b'' coefficient should not be dependent on the state of annealing. The results obtained here for b'' are $-1{\cdot}1$ and $-1{\cdot}3$ respectively for annealed and unannealed cobalt but, in view of the difficulties of measuring the reversible changes, the slight variation in the measured b'' values is not thought to be

significant. From the theory, equation (2), it is suggested that for rotational effects in cobalt b'' is $-1{\cdot}27$, in very good agreement with the experimental values obtained above.

§ 6. CONCLUSION

From the results on cobalt it is clear that the new direct method provides a satisfactory initial separation of the reversible and irreversible thermal components. It is, however, limited to materials in which the incremental susceptibility is nowhere large; but even in less favourable cases it should be of use in giving a general indication of the components.

The method appears to be sound and applicable to any material provided that the thermal sensitivity of the measuring apparatus can be sufficiently increased. Perhaps this could be achieved by using a split photocell amplifier, but such a system could be employed successfully only if there were a marked reduction in the random thermal drift associated with the apparatus.

ACKNOWLEDGMENT

One of us (N.P.R.S.) wishes to thank the Department of Scientific and Industrial Research for a maintenance allowance received during part of this work.

REFERENCES

BATES, L. F., 1951, *J. Phys. Radium*, **12,** 459.

BATES, L. F., and HARRISON, E. G., 1948, *Proc. Phys. Soc.* A, **60,** 213.

BATES, L. F., and SHERRY, N. P. R., 1955, *Proc. Phys. Soc.* B, **68,** in the press.

MEYER, A. J. P., and TAGLAND, P., 1954, *C.R. Acad. Sci., Paris*, **239,** 1611.

OKAMURA, T., 1936, *Sci. Rep. Tôhoku Univ.*, **24,** 744.

RHODES, P., 1948, *Proc. Leeds Phil. Lit. Soc.*, **5,** 116.

STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, **40,** 481.

TEBBLE, R. S., WOOD, J. E., and FLORENTIN, J. J., 1952, *Proc. Phys. Soc.* B, **65,** 858.

The Properties of Fine Virtual Slits formed by Oblique Reflection at Cylindrical Surfaces

By A. H. McILRAITH
Physics Department, Canterbury University College, Christchurch, New Zealand

MS. received 3rd January 1955 *and in revised form* 23*rd March* 1955

Abstract. A theoretical and experimental examination of a virtual slit formed by reflection at a convex circular cylinder suggests that it is well suited for use as an entrance slit in spectroscopy. An expression for the aberration of the reflected wave front is found by making use of the properties of the virtual caustic. Measurements made on the wave front with a Lloyd mirror are in agreement with this expression. For a given tolerance on the aberration, the upper limit to the radius of the cylinder that can be used is inversely proportional to the cube of the angular aperture over which the reflected light is collected, and is from ten to twenty wavelengths for an aperture of 60°. The image is always rectangular; it follows precisely the much greater movements of the source slit and can thus be opened symmetrically even at widths of the order of one wavelength. Its brightness is from 10% to 20% lower than that of the source. Attention is drawn to its use as a curved slit and to its advantages over conventional slits at widths approaching a wavelength.

---

§ 1. Introduction

The investigations of Schuster (1905), van Cittert (1930) and Stockbarger and Burns (1933) have demonstrated the need in spectroscopy for fine entrance slits, ranging between one and twenty wavelengths in width if the highest resolution is to be achieved. It is well known that such narrow slits are very difficult to manufacture especially if they are required to open symmetrically. With this in mind the writer was prompted by an article on the production of fine slits by refraction (King 1946) to explore the properties of the virtual image of a slit formed by reflection at a convex cylindrical surface.

This paper presents the theoretical and experimental results of these investigations with particular reference to the aberration of the reflected wave front, the slit width and the intensity of the final spectral image.

THEORY

§ 2. Aberration of the Reflected Wave Front

Consider first the reflection of a plane wave front by a cylinder of radius r whose axis is parallel to the wave front. The aberration of the wave front from the ideal wave front, a circular cylinder whose axis cuts the axis of the collimator and lies in its focal plane, can be found by the methods of geometrical optics, advantage being taken of the fact that the virtual caustic formed by reflection is the evolute of the reflected wave front.

We shall neglect the dependence of the phase and amplitude of the reflected wave on the angle of incidence ϕ (figure 1 (*a*)). This approximation is sufficient for the present purpose since in practice $30° < \phi < 60°$, so that the phase change on reflection does not vary by more than a few degrees from its mean value. Although the intensity may vary by as much as 30% from its mean value, simple calculations show that this increases the width of the resultant image by less than 0·4%.

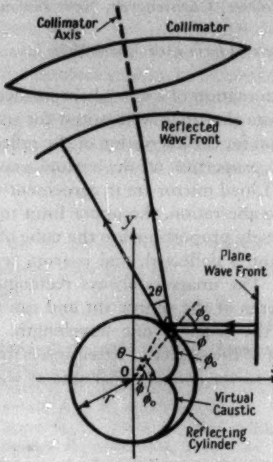

Collimator Axis
Collimator
Reflected Wave Front
y
Plane Wave Front
2θ
φ₀
φ
φ₀
θ
O
x
φ
φ₀
r
Virtual Caustic
Reflecting Cylinder

Figure 1 (*a*).

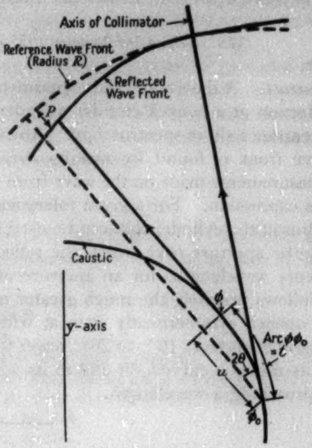

Axis of Collimator
Reference Wave Front (Radius R)
Reflected Wave Front
p
Caustic
y-axis
φ
Arc φφ₀ = t
2θ
u
φ₀

Figure 1 (*b*).

Let the focus of the collimator be at ϕ_0 (figure 1 (*b*)), and let its axis be tangential to the caustic at this point. We take the reference cylinder, from which the errors in wave front are measured, to be centred at ϕ_0. For any other value of ϕ, corresponding to reflection not on the collimator axis, the error p in the reflected wave is, in the notation of figure 1 (*b*), $p = t - u$, $R \gg r$. It is easily shown that the caustic has the form

$$\left.\begin{aligned} x &= \tfrac{1}{2}r(1+2\sin^2\phi)\cos\phi \\ y &= r\sin^3\phi \end{aligned}\right\} \qquad \ldots\ldots(1)$$

then

$$\left.\begin{aligned} t &= \tfrac{3}{2}r(\cos\phi_0 - \cos\phi) \\ u &= \tfrac{1}{2}r(\cos\phi + 2\sin^2\phi\cos\phi - \cos\phi_0 - 2\sin^2\phi_0\cos\phi_0)\cos 2\phi \\ &\qquad + r(\sin^3\phi - \sin^3\phi_0)\sin 2\phi \end{aligned}\right\} \qquad \ldots\ldots(2)$$

substituting $\theta = \phi - \phi_0$

$$p = r\theta^3\sin\phi_0 + \frac{r\theta^4}{4}\cos\phi_0 - \frac{r\theta^5}{4}\sin\phi_0 - \frac{r\theta^6}{24}\cos\phi_0 + \ldots \qquad \ldots\ldots(3)$$

and if only the leading term is retained

$$p = r\theta^3\sin\phi_0 + O(\theta^4). \qquad \ldots\ldots(3a)$$

The most serious effect on a spectral image of aberration having this form is perhaps the reduction in contrast caused by the increase in the intensity I_1 of one of the subsidiary maxima relative to the intensity I_0 of the principal maximum (Rayleigh 1899). As the ratio I_1/I_0 changes from 0·047 when the maximum aberration $P_{\max}=0$ to 0·07 when $P_{\max}=\lambda/8$ and to 0·137 when $P_{\max}=\lambda/4$, this effect is not serious until the lines to be resolved differ by about a factor of five in intensity. Then it would seem desirable to use the criterion $P_{\max}\leqslant\lambda/8$, otherwise the criterion $P_{\max}\leqslant\lambda/4$ should suffice. To satisfy these conditions

$$\left.\begin{array}{ll} & r/\lambda\leqslant 1/8\theta^3{}_{\max}\sin\phi_0=8/\Omega^3\sin\phi_0 \\ \text{or} & r/\lambda\leqslant 1/4\theta^3{}_{\max}\sin\phi_0=16/\Omega^3\sin\phi_0 \end{array}\right\}\quad\ldots\ldots(4)$$

where Ω is the angular aperture of the collimator.

Provided $2\pi r/\lambda\gg 1$ the results obtained by physical and geometrical optics are almost coincident (Morse and Feshbach 1953). Therefore equation (4) can safely be used to find the maximum permissible value of r/λ for systems with apertures up to 60°.

When the point source is a finite distance a from the cylinder, the corresponding caustic is a surface lying parallel to the axis of the cylinder and with a cross section similar to that described above. Provided $a\geqslant 4r$, the aberration is given to a good approximation by $p\sin\gamma$, where p is given by equation (3a) and γ is the angle between the axis of the cylinder and the normal to the wave front.

§ 3. WIDTH OF THE VIRTUAL SLIT

If light waves from two distant sources, Q_0 and Q_1 (figure 2), strike the cylinder the reflected wave fronts appear to diverge from the points V_0 and V_1 where tangents passing through the centre of the collimator touch the corresponding caustics, i.e. V_0 and V_1 are the Gauss images of Q_0 and Q_1.

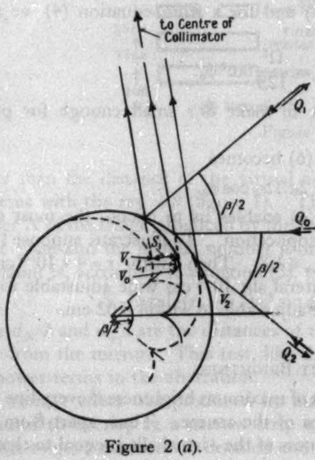


Figure 2 (a).

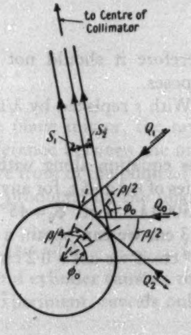


Figure 2 (b).

Resolving the distance V_0V_1 into orthogonal components, s_1 and l_1 (figure 2), we have

$$s_1 = 2r\sin(\beta/8)\cos(\phi_0 - \beta/8)$$

similarly $$s_2 = 2r\sin(\beta/8)\cos(\phi_0 + \beta/8).$$

Therefore, if Q_1 and Q_2 are at the opposite edges of a slit source subtending an angle β at the cylinder the total width of the image resolved perpendicular to the collimator axis is

$$s = s_1 + s_2 = 4r\sin(\beta/8)\cos\phi_0\cos(\beta/8) \simeq \tfrac{1}{2}r\beta\cos\phi_0. \qquad \ldots\ldots(5)$$

If the source slit of width s_0 is at a distance a from the cylinder equation (5) becomes

$$s \simeq (rs_0/2a)\cos\phi_0. \qquad \ldots\ldots(6)$$

By making use of equation (1) and the fact that caustics 1 and 2 are rotated through angles $\pm\beta/2$ with respect to caustic O we find

$$l = l_1 + l_2 \simeq \tfrac{1}{2}s\tan\phi_0. \qquad \ldots\ldots(7)$$

Since the image acts as an incoherently lit slit, the optimum value of s is λ/Ω, any further increase in s producing an appreciable broadening of the spectral image with only a slight increase in its intensity.

Using this value for s and noting that the depth of field δf of the collimator is given by $\delta f = 4\lambda/\Omega^2$, we find that l is less than $\delta f/8$ for $\phi_0 \leqslant 45°$ and $\Omega \leqslant 60°$ and consequently causes no significant defocusing. As the image of each point source is a line parallel to the axis of the cylinder, the image of the source slit is rectangular.

The displacement δs of the centre of the image from the image of the centre of the source slit is

$$\delta s = (s_1 - s_2)/2 = 2r\sin\phi_0\sin^2(\beta/8)$$

and $$\delta s/s = \tfrac{1}{2}\tan\phi_0\tan(\beta/8) \simeq \tfrac{1}{16}\beta\tan\phi_0.$$

Substituting for β from equation (5) and for r from equation (4) we find on putting $s = \lambda/\Omega$ that

$$(\delta s/s)_{\min} = \frac{\Omega^2}{128}\tan^2\phi_0. \qquad \ldots\ldots(8)$$

Therefore it should not be difficult to make $\delta s/s$ small enough for practical purposes.

With s replaced by λ/Ω equation (6) becomes

$$r/\lambda = 2a/\Omega s_0\cos\phi_0. \qquad \ldots\ldots(9)$$

This equation along with equation (4) enables us to choose the most suitable values of r, a and s_0 for any particular application. To illustrate, suppose $\Omega = 30°$, $\lambda = 5000$ Å. Let $\phi_0 = 45°$ and $P_{\max} \leqslant \lambda/8$. Then with $r = 4 \times 10^{-3}$ cm and $a = 3$ cm we can obtain a virtual bilateral slit 10^{-4} cm wide adjustable to within 10^{-5} cm if we use a 0·2 cm source slit adjustable to within 0·02 cm.

§4. SLIT BRIGHTNESS

As with a real slit, the virtual slit is of maximum brightness if every line passing through it intersects an emitting area of the source. Thus, apart from a 10% to 20% loss on reflection, the brightness of the virtual slit is equal to that of the source if the source is of height not less than $a\Omega + h(a+f)/f$ where f is the focal length of the collimator and h is the height utilized of the virtual slit. This compares very favourably with the performance of real slits as used in practice.

It should be noted that when $s = \lambda/\Omega$,

$$\beta/\Omega = 2s/\Omega r \cos\phi_0 = 2\lambda/r\Omega^2 \cos\phi_0 \geqslant \tfrac{1}{4}\Omega \tan\phi_0$$

according to equations (4) and (5). Consequently the angle β subtended by the source slit at the cylinder can be made much smaller than the angle Ω subtended by the collimator. This angle transforming property is of particular advantage when the source employed is so narrow that in normal practice it would be necessary to use a condenser lens to fill the collimator with light.

EXPERIMENT

§5. Aberration of the Reflected Wave Front

A Lloyd mirror experiment as illustrated in figure 3 was performed to test the validity of equation (3). Two sets of fringes were recorded on the photographic plate, the first due to wave fronts reflected from an aluminized glass rod (0·36 cm diameter), the second due to wave fronts reflected from a nickel-plated sewing needle (0·068 cm in diameter). The needle served instead of a fine slit as the source of the reference wave front as none of the slits available were able to produce fringes over a sufficiently large angle (20°) for an adequate test of this theory to be made. With the distance of the plate from the cylinder much greater than the distance of the virtual slit from the plane mirror, the ray $-\theta$ interferes with the ray $+\theta$ (figure 1). Then the difference between the ordinal number N of the fringe produced by the reference wave front ot the point $\omega\,(=2\theta)$ on the plate and the ordinal number n of the fringe produced by the unknown wave front is, according to equation (3), given by

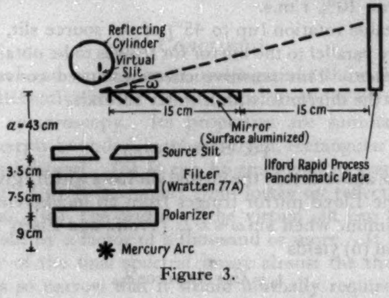


Figure 3.

$$N - n \simeq [\tfrac{1}{4}(r_n - r_N)\omega^3 \sin\phi_0 + (d_N - d_n)\sin\omega]/\lambda \qquad \ldots\ldots(10)$$

where $d_N/2$ and $d_n/2$ are the distances of the needle and cylinder caustics respectively from the mirror. This test, like the Young experiment, reveals only the odd power terms in the aberration.

A vernier effect was produced by making $d_N \simeq 1{\cdot}1 d_n$ and allowing the two exposures to overlap in a narrow strip of the plate (see Plate). This made it possible to get a much more detailed and accurate estimate of the dependence of

$N-n$ on N and hence on ω. Since every cycle of the modulated fringes marks a change of unity in $N-n$, $N-n$ can be found directly as a function of N. The data thus obtained were initially expressed by the method of least squares as a power series in N terminated at N^3, the coefficients of N^4 and N^5 being less than twice their standard deviations. In addition N was found by means of a travelling microscope as a function of position on the plate. It was then expressed as a power series in ω by the above method.

Substitution of this expression for N into the expression for $N-n$ yields

$$N-n=-(2\cdot783\pm0\cdot031)+(523\cdot9\pm1\cdot0)\omega-(7\pm8)\omega^2+(450\pm19)\omega^3, \quad \ldots\ldots(11)$$

the limits quoted being the estimated standard deviations. The small standard deviation, 0·040 fringe, of the experimental points from this regression line reflects the artificial sharpening of the fringes due to the use of a photographic plate of high contrast.

The expected expression for $N-n$ is

$$N-n=-(3\pm0\cdot4)+(390\pm200)\omega-(6\cdot5\pm4\cdot6)\omega^2+(420\pm30)\omega^3. \quad \ldots\ldots(12)$$

This is based on equation (10) and measurements of the cylinder diameter etc. Allowance is made for a small unexpected discrepancy in the expression for N. This effect, assumed to be due to curvature of the plane mirror, accounts for $-6\cdot5\omega^2+30\omega^3$ in equation (12). The constant term and that in ω are due to a zero error in ω and the factor d_N-d_n in equation (10) respectively; the remainder of the term in ω^3 is given by equation (10). As the contribution $\frac{1}{4}(r_n-r_N)\omega^3\sin\phi_0$ amounts to $477\omega^3$ it follows that the first term in equation (3*a*) agrees with experiment to within 10% r.m.s.

Through a suitable rotation (up to 45°) of the source slit, the virtual slit can be made sufficiently parallel to the mirror for fringes to be obtained over the whole region of interference. This sensitive effect was used conversely to make the cylinder parallel to the mirror to within thirty seconds.

§6. SLIT WIDTH

According to the above theory the virtual slit has a width s given by equation (6). The visibility of the Lloyd mirror fringes from an incoherently lit slit of width s has its first minimum when $\sin\omega=\lambda/2s$ (Wolfe and Eisen 1948). Then substitution in equation (6) yields

$$s_0=\lambda a/r\sin\omega\cos\phi_0. \quad \ldots\ldots(13)$$

This equation was tested by the experimental arrangement shown in figure 3 with a reduced to 10 cm and the photographic plate replaced by an eyepiece.

r (cm)	ω	s_0 (10^{-3} cm)	
		Measured	from eqn. (13)
0·510	16° 40′	7·5	5·3
	8° 30′	12·3	10·3
0·184	16° 40′	15·3	14·8
	8° 30′	30·3	28·6

The results presented in the table show that the directly measured values of s_0 are from 4% to 45% greater than those derived from fringe visibility thus indicating that the reflected light was *more* coherent than expected. As the errors involved in the determination of the quantities in equation (13) are negligible, the most probable explanation of these discrepancies is that, as the 1 mm wide source was placed 20 cm behind the source slit, the latter was more coherently lit than is assumed in the theory. As expected, this discrepancy is less for the smaller cylinder and does not exceed 6%. Hence equation (6) is sufficiently accurate for most applications.

§ 7. MANUFACTURE OF CYLINDERS OF SMALL RADIUS

Experience gained in this investigation suggests that aluminized cylinders of glass or quartz drawn out from rod may provide a simple solution to this problem, as they should prove satisfactory with respect to smoothness and perfection of figure at the smaller diameters demanded by equation (4). The satisfactory performance of steel sewing needles shows that they form an alternative source of supply when $r > 0{\cdot}01$ cm, i.e. $\Omega < 22°$, provided they can be held sufficiently straight.

Fine wires under tension may be useful when Ω approaches 60°, i.e. r approaches 10λ, although microscopic examination of a few samples suggests that it may be difficult to find wire with a sufficiently smooth surface. A slight tapering merely results in a proportionate tapering of the virtual slit.

It should be possible to keep the cylinder sufficiently straight to make the resultant curvature of the spectral lines negligible. Bowing is more serious in the Lloyd mirror experiment where it reduces the fringe visibility and height. It is revealed by fringe movement parallel to the mirror when the source slit is rotated.

§ 8. CONCLUSION

These theoretical and experimental investigations show that this virtual slit offers a possible alternative to the conventional methods of manufacturing fine entrance slits for spectroscopy. Its properties are summarized as follows :

(i) It acts as a sharply bounded, incoherently lit, rectangular slit. It is bilateral if the source slit is bilateral. (ii) It can be employed in systems using angular apertures up to 60°. (iii) It has, apart from losses on reflection, the maximum possible brightness. (iv) The width of the virtual slit can be made less than that of the source slit by a factor of a thousand or more. It is often possible to make the intensity of the final spectral image almost the theoretical maximum when the source is so narrow that it would normally require a condenser lens. This applies also to sources extended in depth as in Raman spectroscopy. (v) Reflecting cylinders appear to have particular application in the manufacture of curved bilateral entrance slits as it should be a simple matter to bend them to the required curvature.

It is of particular interest to compare the performance of this type of slit with that of a conventional slit when the slit width is of the order of one wavelength.

The 'transmission' and polarizing properties of the virtual slit are nearly independent of r and s for $r \geqslant 10\lambda$ and $s \geqslant \lambda$. Its transmitted wave front is asymmetrical in intensity, the relative value of the latter varying from 1·25 at $\phi = 30°$ to 1·00 at $\phi = 45°$ and to 0·72 at $\phi = 60°$. With an aluminium surface the state of polarization changes from 1·5% at $\phi = 30°$ to 7% at $\phi = 60°$, the 'transmission' being greater for light polarized parallel to the slit (Schulz 1954, Schulz and Tangherlini 1954).

Jones and Richards (1954) have shown that the transmission, polarizing and diffracting properties of a typical spectroscopic slit depend markedly on its depth and electrical conductivity as well as on its width. With a depth of 35μ the transmission is almost zero at a width of one wavelength while the transmitted light is strongly polarized, the component polarized parallel to the slit predominating. Finally, as the diffraction pattern departs from that due to a slit of zero depth when the slit width falls below 100 wavelengths, it is possible that the transmitted wave front suffers from appreciable aberration.

ACKNOWLEDGMENTS

I wish to acknowledge my indebtedness to the late Professor F. C. Chalklin for encouragement and support received in this work. I also wish to thank Dr. G. T. P. Tarrant and Dr. N. K. Pope of Canterbury University College and Dr. G. L. Rogers of Victoria University College for their advice during the writing of this article, and Mr. E. Mangin for advice and assistance in photography. Thanks are also due to Imperial Chemical Industries, New Zealand, Ltd., and the University of New Zealand for the award of Research Fellowships, and to the University Grants Committee for the provision of equipment.

REFERENCES

JONES, R. V., and RICHARDS, J. C. S., 1954, *Proc. Roy. Soc.* A, **225,** 122.
KING, C. M., 1946, *J. Opt. Soc. Amer.*, **36,** 164.
MORSE, P. M., and FESHBACH, H., 1953, *Methods of Theoretical Physics* (New York: McGraw-Hill), pp. 1380–1382.
RAYLEIGH, LORD, 1899, *Scientific Papers*, Vol. 1 (Cambridge: University Press), p. 430.
SCHULZ, L. G., 1954, *J. Opt. Soc. Amer.*, **44,** 357.
SCHULZ, L. G., and TANGHERLINI, F. R., 1954, *J. Opt. Soc. Amer.*, **44,** 362.
SCHUSTER, A., 1905, *Astrophys. J.*, **21,** 197.
STOCKBARGER, D. C., and BURNS, L., 1933, *J. Opt. Soc. Amer.*, **23,** 379.
VAN CITTERT, P. H., 1930, *Z. Phys.*, **65,** 547.
WOLFE, R. N., and EISEN, F. C., 1948, *J. Opt. Soc. Amer.*, **38,** 706.

Theory of Secondary Electron Emission from Solids

By H. FRÖHLICH
Department of Theoretical Physics, The University of Liverpool

MS. received 1st *April* 1955

Abstract. A formula is derived permitting the calculation of the yield of secondary electrons from the yield of photoelectrons.

§ 1. Introduction and Discussion

The emission of secondary electrons from solids is a very involved process, largely in view of internal scattering of these electrons before their emission. In the present paper it is intended to derive a correlation between photoelectric and secondary emission which in favourable circumstances does not require knowledge of internal scattering. For this purpose the primary particle will be replaced by its equivalent radiation field (Fermi 1924, Weizsäcker 1935, Williams 1935) whose yield of photoelectrons can be obtained from photoelectric measurements. This method has its limitations. It would not work for completely free (non-interacting) electrons which would not show any photoelectric emission. However, although for certain purposes electrons in metals behave like free ones with a certain effective mass, the actual electronic wave functions of the Bloch type

$$\psi = \exp(i\mathbf{kr})u_{\mathbf{k}}(\mathbf{r}) \qquad \ldots\ldots(1.1)$$

differ by the periodic function (lattice period) $u_{\mathbf{k}}(\mathbf{r})$ from free electron wave functions and in general $u_{\mathbf{k}}(\mathbf{r})$ cannot be approximated by a constant. This holds for both metals and insulators. It is strange, therefore, to notice that recent papers on this subject (for a summary see the report by Hachenberger and Brauer (1954)) make just this assumption thus completely disregarding the knowledge of wave functions in solids accumulated during the last two decades. In fact the arguments on which these papers are based are equivalent to assuming that the oscillator strengths for all optical transitions are negligibly small (they vanish for free electrons) which is wrong for nearly all solids.†

The method proposed here is analogous to the method of calculating the energy loss of a fast particle in a solid in terms of its (frequency dependent) dielectric constant (cf. Fröhlich and Platzman 1953). There, the primary particle is replaced by a moving point charge and its energy loss and scattering are neglected to the first order. The electric displacement of the moving point charge is developed in a Fourier series with respect to time and the energy transfer to the solid due to each frequency can then be expressed in terms of a complex dielectric constant $\epsilon(\omega)$. For each frequency ω the field also depends on the distance from the point charge. The corresponding spatial Fourier development would permit a dependence of the dielectric constant on wavelength λ as well as on frequency $(\epsilon(\omega, \lambda))$. Usually this is replaced by a step function, $\epsilon(\omega, \lambda) = \epsilon(\omega)$ if $\lambda > \lambda_0$ and $\epsilon(\omega, \lambda) = 1$ if $\lambda < \lambda_0$. The cut-off λ_0 has to be determined from independent arguments; the total energy loss, however, depends only logarithmically on λ_0 so that the resulting energy loss does not sensitively depend on λ_0. The value of $\epsilon(\omega)$ may be taken either from experiment or from theory.

---

† Another serious mistake occurring in some of these papers is to assume that the screening radius of the primary electron in a metal is a constant independent of its velocity; this is quite wrong in the range mainly of interest between say 50 ev and several thousand ev.

In the present paper it is proposed to neglect again energy loss and scattering of the primary particle to the first order, and to develop its electric displacement in a Fourier series in time. To each frequency can then be coordinated a photo-yield, either empirically or theoretically. In the former case internal scattering of the secondary electrons is automatically taken into account. Again a spatial cut-off will have to be introduced as in the calculation of energy loss. The method applies in the first place when the following two conditions are fulfilled: (i) most secondary electrons come from a layer of thickness l within which the primary particle has not appreciably changed its velocity; (ii) the penetration depth of light of the most important frequencies in the equivalent radiation field is larger than l. These conditions can always be fulfilled if the experiment on secondary electron emission is carried out on sufficiently thin layers. In favourable circumstances they may also hold for bulk materials. If this is not the case then the emission from bulk materials may in principle be obtained by adding the contributions of individual layers of thickness l; the absorption of both primaries and secondaries in these layers must then be known, however. This case will not be treated here.

Clearly the most favourable conditions occur in solids with very thin but highly photosensitive surface layers. For then, if the equivalent radiation field has an appreciable intensity in the frequency region which is highly photosensitive, emission from lower layers can be neglected so that both the above conditions are fulfilled. From the above arguments it should be expected then that such layers also show a high secondary electron yield as has actually been observed (cf. Bruining 1954). The following formulae will give the possibility of a quantitative check.

Subject to the validity of the method in general the following developments hold for both metals and insulators. Since wave functions for individual electrons are qualitatively the same in both cases, the only appreciable difference is in the screening of the primaries. In the present treatment this is automatically taken into account by expressing the equivalent radiation field in terms of the electric displacement. Thus if the photo-yield is known in terms of the intensity of the incoming light it must first be expressed in terms of the magnitude of the electric displacement. This requires introduction of the dielectric constant $\epsilon(\omega)$ which is a measure for the screening of a charge.

Thus for a detailed experimental check of the theory both photoelectric and secondary electron yield should be measured on the same sample (a thin layer or a solid with a photosensitive surface layer). Furthermore the optical constants in the photosensitive frequency region must be determined so as to evaluate the electric displacement from the light intensity, and hence the functions $f_l(\mathbf{v}, \omega)$ according to (2.11). Use of equations (2.12) or (2.13) then leads to the yield of secondary electrons in terms of the photo-yield.

§ 2. CALCULATIONS

Let $\mathbf{d}(p,t)$ be the electric displacement due to a moving point charge e (velocity $\mathbf{v}_0$) at time t at a point P whose shortest distance (collision parameter) from the path $\mathbf{v}_0$ is p. Then its components perpendicular and parallel to $\mathbf{v}_0$ respectively are

$$d_\perp(p,t)=\frac{ep}{[p^2+v_0^2(t-t_0)^2]^{3/2}},\quad d_\|(p,t)=\frac{ev_0t}{[p^2+v_0^2(t-t_0)^2]^{3/2}},\qquad \ldots\ldots(2.1)$$

where t_0 is the time at which the point charge has the distance p from P. Let then

$$\left.\begin{aligned}\mathbf{d}(p,t)&=\int_{-\infty}^{\infty}\mathbf{d}_\omega(p)\exp\{-i\omega(t-t_0)\}\,d\omega,\\ \mathbf{d}_\omega(p)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{d}(p,t)\exp\{i\omega(t-t_0)\}\,dt,\end{aligned}\right\}\qquad\ldots\ldots(2.2)$$

where reality requires that

$$\mathbf{d}_\omega(p)=\mathbf{d}_{-\omega}{}^*(p).\qquad\ldots\ldots(2.3)$$

The components of $\mathbf{d}_\omega(p)$ corresponding to $d_\perp(p,t)$ and $d_\|(p,t)$ will be denoted by $d_{\perp\omega}(p)$ and $d_{\|\omega}(p)$ respectively. Then if primary particles with velocity $\mathbf{v}_0$ cross the surface at a rate n per second and per unit surface, an average electric displacement $\mathbf{D}$ is established which can be decomposed into its Fourier components. Clearly

$$\begin{aligned}D_\|^2&=n\int_{-\infty}^{\infty}dt_0\int_{p_{\min}}^{\infty}2\pi p\,dp\,d_\|^2(p,t)\\ &=n\int_{-\infty}^{\infty}dt_0\int_{p_{\min}}^{\infty}2\pi p\,dp\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\mu\,d_{\|\omega}(p)\,d_{\|\mu}(p)\exp\{-i(\omega+\mu)(t-t_0)\}\\ &=n(2\pi)^2\int_{-\infty}^{\infty}d\omega\int_{p_{\min}}^{\infty}p\,dp\,d_{\|\omega}\,d_{\|-\omega}=2n(2\pi)^2\int_0^{\infty}d\omega\int_{p_{\min}}^{\infty}p\,dp\,|\,d_{\|\omega}\,|^2\end{aligned}\qquad\ldots\ldots(2.4)$$

and similarly for $D_\perp^2$. Thus if $D_{\|\omega}{}^2\,d\omega$ denotes the square of the electric displacement parallel to the direction of $\mathbf{v}_0$ in a frequency range $d\omega$ near ω, then from (2.4) using (2.1) and (2.2),

$$\begin{aligned}D_{\|\omega}{}^2\,d\omega&=8\pi^2n\,d\omega\int_{p_{\min}}^{\infty}\frac{e^2\omega^2}{\pi^2v_0^4}\mathrm{K}_1^2(\omega p/v_0)p\,dp\\ &=8\frac{e^2}{v_0^2}n\,d\omega\frac{x^2}{2}\{\mathrm{K}_0^2(x)+\mathrm{K}_1^2(x)\}\end{aligned}\qquad\ldots\ldots(2.5)$$

where

$$x=\omega p_{\min}/v_0.\qquad\ldots\ldots(2.6)$$

K_0 and K_1 are the modified Bessel functions of second kind, order 0 and 1 respectively. Similarly one finds for the electric displacement perpendicular to $\mathbf{v}_0$

$$\begin{aligned}D_{\perp\omega}{}^2\,d\omega&=8\pi^2n\,d\omega\int_{p_{\min}}^{\infty}\frac{e^2\omega^2}{\pi^2v_0^4}\mathrm{K}_0^2(\omega p/v_0)p\,dp\\ &=\frac{8e^2}{v_0^2}n\,d\omega\left\{x\mathrm{K}_0(x)\mathrm{K}_1(x)-\frac{x^2}{2}(\mathrm{K}_0^2(x)+\mathrm{K}_1^2(x))\right\}.\end{aligned}\qquad\ldots\ldots(2.7)$$

It should be noted that $D_\omega{}^2\,d\omega$, and not $D_\omega{}^2$, has the dimension of the square of an electric displacement. The smallest value that p can take is denoted by $p_{\min}$. Fortunately the above expressions (2.5) and (2.7) are not very sensitive to the exact value of $p_{\min}$ provided x is sufficiently smaller than unity. In our case $p_{\min}$ should be chosen of the order of magnitude of interatomic distance. The principal frequencies ω required below have an order of magnitude such that $\hbar\omega \sim 10$ ev. If then v_0 is equal to the velocity of an electron with an energy of several hundred ev (or more), $x \ll 1$ is fulfilled. Then since

$$\mathrm{K}_0(x) \simeq \log(2/\gamma x), \quad x\mathrm{K}_1(x) \simeq 1, \quad \gamma = 1{\cdot}78\ldots, \text{ if } x \ll 1, \qquad \ldots\ldots(2.8)$$

approximately,

$$D_{\perp\omega}{}^2\, d\omega \simeq \frac{8e^2}{v_0{}^2} n\, d\omega \left(\log\frac{2}{\gamma x} - \tfrac{1}{2}\right), \text{ if } x \ll 1 \qquad \ldots\ldots(2.9)$$

and

$$D_{\|\omega}{}^2\, d\omega \simeq \frac{1}{2}\frac{8e^2}{v_0{}^2}\, n\, d\omega, \qquad \text{if} \quad x \ll 1. \qquad \ldots\ldots(2.10)$$

Define now a quantity $P_{\mathrm{i}}(\mathbf{v}, \omega)$, $\mathrm{i} = \perp$ or $\|$, such that $P_{\mathrm{i}}(\mathbf{v}, \omega)\, dv_x\, dv_y\, dv_z$ is the rate of emission of photoelectrons with velocity $\mathbf{v}$ in a range $dv_x\, dv_y\, dv_z$ if the square of the electric displacement in the solid (due to light) is $D_{\mathrm{i}\omega}{}^2\, d\omega$ in a range $d\omega$ near ω. It should be noted that knowledge of the optical constants is required to obtain this quantity from the intensity of the incoming light. Then two empirical functions $f_{\mathrm{i}}(\mathbf{v}, \omega)$, $\mathrm{i} = \perp$ or $\|$, can be found such that

$$P_{\mathrm{i}}(\mathbf{v}, \omega) = f_{\mathrm{i}}(\mathbf{v}, \omega) D_{\mathrm{i}\omega}{}^2, \qquad \mathrm{i} = \perp \text{ or } \|. \qquad \ldots\ldots(2.11)$$

Hence if $Q(\mathbf{v})\, dv_x\, dv_y\, dv_z$ is the rate of emission of secondary electrons of velocity $\mathbf{v}$ in a range $dv_x\, dv_y\, dv_z$ due to primaries of velocity $\mathbf{v}_0$ crossing the surface at a rate n per second per unit surface then with (2.5) and (2.7)

$$\begin{aligned} Q(\mathbf{v}) &= \int f_\perp(\mathbf{v}, \omega) D_{\perp\omega}{}^2\, d\omega + \int f_\|(\mathbf{v}, \omega) D_{\|\omega}{}^2\, d\omega \\ &= \frac{8e^2}{v_0{}^2} n \int \left\{ x\mathrm{K}_0(x)\mathrm{K}_1(x) f_\perp(\mathbf{v}, \omega) + \frac{x^2}{2}(\mathrm{K}_0{}^2(x) + \mathrm{K}_1{}^2(x))(f_\|(\mathbf{v}, \omega) - f_\perp(\mathbf{v}, \omega)) \right\} d\omega. \end{aligned} \qquad \ldots\ldots(2.12)$$

If in particular $x \ll 1$ which should always hold for sufficiently large velocity v_0 of the primaries, use of formulae (2.8)–(2.10) gives

$$Q(\mathbf{v}) \simeq \frac{8e^2}{v_0{}^2} n \int \left\{ f_\perp(\mathbf{v}, \omega) \log\frac{1{\cdot}1 v_0}{\omega p_{\min}} + \frac{1}{2}(f_\|(\mathbf{v}, \omega) - f_\perp(\mathbf{v}, \omega)) \right\} d\omega, \text{ if } \frac{\omega p_{\min}}{v_0} \ll 1. \qquad \ldots\ldots(2.13)$$

It should be remembered that $\perp$ and $\|$ refer to the direction of the incoming beam, not to the surface. It will be noticed that the expressions simplify if $f_\perp = f_\|$.

Clearly, (2.13) in the range in which it holds, is not very sensitive to the exact value of $p_{\min}$. However in the region in which $Q(\mathbf{v})$ has its maximum as a function of v_0, equation (2.12) must be used, and the value of v_0 at which the maximum occurs may be fairly sensitive to the exact value of $p_{\min}$. It might then be preferable to determine $p_{\min}$ from the position of the maximum.

REFERENCES

BRUINING, H., 1954, *Physics and Applications of Secondary Electron Emission* (London : Pergamon Press).

FERMI, E., 1924, *Z. Phys.*, **29**, 315.

FRÖHLICH, H., and PLATZMAN, R. L., 1953, *Phys. Rev.*, **92**, 1152.

HACHENBERGER, O., and BRAUER, W., 1954, *Fortschr. Phys.*, **1**, 439.

v. WEIZSÄCKER, C. F., 1935, *Z. Phys.*, **88**, 612.

WILLIAMS, E. J., 1935, *K. Danske Vidensk. Selsk. Mat. Fys. Medd.*, **13**, No. 4.

Fluctuations of the Numbers of Electrons and Holes in a Semiconductor

By R. E. BURGESS†
Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose; MS. received 24th September 1954 and in amended form 11th March 1955

Abstract. In calculations of certain types of noise in semiconductors it is required to know the fluctuation in the number of mobile charge carriers (electrons and holes). These fluctuations are calculated for a non-degenerate semiconductor in thermal equilibrium by two different methods: the first in terms of the electronic free energy of the system whose minimal properties are used to determine the equilibrium distribution and fluctuations of electrons between the conduction and valence bands and the impurity levels. The second approach is purely statistical and is based on the probabilities of the elementary processes of electron transitions between energy levels. The results of the two methods are identical but the thermodynamical method yields the solutions for complex cases which are not readily dealt with by the statistical method. Both intrinsic and extrinsic semiconductors are discussed, as well as the important practical case of a nearly intrinsic semiconductor.

§ 1. INTRODUCTION

In considering the fluctuations of current in a semiconductor to which a steady field is applied it is required to know the extent of the fluctuations in the numbers of mobile charge carriers (i.e. electrons in the conduction band and holes in the valence band) due to the randomness in the process of generation and recombination.

If the semiconductor is intrinsic, i.e. electrons can only arise from the atoms of the semiconductor in a perfect lattice (so that there are no impurity or defect levels), then the numbers of electrons and holes must be equal at any given moment. If the semiconductor is strongly extrinsic and is say n-type due to a relatively large density of donor impurity atoms, most of the electrons in the conduction band will have been excited from the donor atoms and a negligible number from the semiconductor atoms: electrostatic interaction between the ions and the carriers ensures (by virtue of Poisson's equation) that neutrality very closely represents the conditions obtaining in any region of the semiconductor of linear dimensions greater than about 10^{-4} cm except in transition regions, near electrodes or at the surface.

In the present paper the following problem will be discussed: the determination of the mean and mean square value of the number of electrons n in the conduction band of a volume V of homogeneous non-degenerate semiconductor in thermodynamical equilibrium at absolute temperature T.

Two types of treatment are adopted. The first is thermodynamical and based on the free energy of the electrons in the conduction and valence bands and of those bound on the impurity levels; the condition of minimum free energy gives the equilibrium condition while the departure of the free energy from this minimum value in terms of the fluctuations in the distribution of electrons over the energy levels enables these fluctuations to be determined. The second approach is purely statistical and from the probabilities of the elementary processes of transitions of an electron from one level to another, the distribution function of the number of electrons in the conduction band is calculated. From this function the mean and the mean square values are formed and the resulting fluctuations are shown to be consistent with those calculated thermodynamically.

---

† Now at the Physics Department, University of British Columbia, Vancouver.

The link between the two methods is the law of mass action applied to the 'reactions' corresponding to the thermal creation of an electron and a positive ion or a hole, balanced by their reversible recombination to form a neutral atom. The equilibrium constant for such reactions can be expressed either in terms of the thermodynamical parameters of the reaction or in terms of the ratio of the probabilities of generation and recombination of an electron (or hole).

In some of the literature on semiconductor noise (e.g. van der Ziel 1953 a, b) erroneous expressions are used for the fluctuations in the number of electrons and holes and these appear to be based on a faulty application of probability analysis. It is therefore instructive to derive the fluctuations by two independent methods having differing fields of generality. The thermodynamical treatment applies to those cases in which thermal equilibrium is established while the statistical approach is capable of application to systems in which equilibrium does not exist, assuming that the transition probabilities for electrons in these systems can be evaluated.

§ 2. THERMODYNAMICAL TREATMENT

2.1. *The Electronic Free Energy and its Derivatives*

The minimum work ΔW which must be done on a system to take it from its equilibrium state to some other state is equal to the change ΔF of the Helmholtz free energy F between the two states. In equilibrium at constant volume and temperature the free energy is a minimum. In general the free energy is defined as $F = U - TS$ where U is the internal energy of the system, T is the absolute temperature and S is the entropy and it will always be of the form $nf(n/V, T)$ where n is the number of particles in the system. If we wish to evaluate the fluctuating quantities x_r then we write $\delta F = 0$ for small changes in x_r:

$$\sum_r \frac{\partial F}{\partial x_r} \Delta x_r = 0.$$

A second order effect due to quadratic terms is related to the fluctuations which we wish to calculate since

$$(\Delta W)_{T,V} = \Delta F = \tfrac{1}{2} \sum_r \sum_s \frac{\partial^2 F}{\partial x_r \partial x_s} \Delta x_r \Delta x_s$$

will have a distribution function of the form $\exp(-\Delta F/\boldsymbol{k}T)$ in classical statistics where $\boldsymbol{k}$ is Boltzmann's constant and thus the various Δx_r have a joint multivariate gaussian distribution.

If now the variables x_r are identified with the number of electrons n_r occupying the various levels, the condition for equilibrium becomes $\Sigma_r \mu_r \Delta n_r = 0$ where $\mu_r \equiv \partial F/\partial n_r$. For this relation to hold identically, all the μ_r must be equal since we have $\Sigma\Delta n_r=0$ in a system where the total number of electrons is conserved. If there are a_r levels of energy ϵ_r the number of ways of arranging μ_r electrons among them is

$$W_r=\frac{a_r\,!}{(a_r-n_r)\,!\,n_r\,!}\,2^{\nu_r}$$

where 2^{ν_r} represents the number of possible arrangements arising from spin.

Using Stirling's approximation for the factorials in W_r we find

$$F=\sum[n_r\epsilon_r-\boldsymbol{k}T\{\nu_r\ln 2+a_r\ln a_r-n_r\ln n_r-(a_r-n_r)\ln(a_r-n_r)\}]$$

and hence

$$\mu_r\equiv\frac{\partial F}{\partial n_r}=\epsilon_r-\frac{\partial\nu_r}{\partial n_r}\boldsymbol{k}T\ln 2+\boldsymbol{k}T\ln\frac{n_r}{a_r-n_r}.$$

This quantity is the electronic electrochemical potential associated with the rth level and in equilibrium all the μ_r must be equal (to μ, say). Thus the equilibrium distribution of n_r is

$$n_r=\frac{a_r}{1+\alpha_r\exp(\epsilon_r-\mu)/\boldsymbol{k}T}\equiv f_r a_r$$

where $\alpha_r\equiv 2^{-\partial\nu_r/\partial n_r}$ and f_r is the fractional occupancy of the rth level by electrons. The second partial derivatives of F with respect to n_r are

$$\frac{\partial^2 F}{\partial n_r^{\,2}}=-\frac{\partial^2\nu_r}{\partial n_r^{\,2}}\boldsymbol{k}T\ln 2+\boldsymbol{k}T\left[\frac{1}{n_r}+\frac{1}{a_r-n_r}\right]$$

where the first term usually vanishes since ν_r is normally a linear function of n_r.

2.2. *Application to Semiconductors*

In a volume V of n-type semiconductor in thermal equilibrium at absolute temperature T there will exist a distribution of electrons between the valence band, the conduction band and the bound levels lying between them. This distribution and the fluctuations in it are readily calculated in terms of the electronic free energy F of the system which may be regarded as the sum of the three components associated with the valence and conduction bands and the donor levels:

$$F=F_{\mathrm{V}}+F_{\mathrm{C}}+F_{\mathrm{D}}. \qquad \ldots\ldots(1)$$

For simplicity a non-degenerate semiconductor will be considered. Now the free energy associated with the conduction band containing n electrons obeying classical statistics is (see, for example, Mott and Gurney 1948)

$$F_{\mathrm{C}}=-n\boldsymbol{k}T\left[\frac{3}{2}\ln\frac{2\pi m_{\mathrm{n}}\boldsymbol{k}T}{\boldsymbol{h}^2}+1+\ln\frac{2V}{n}\right] \qquad \ldots\ldots(2)$$

where m_{n} is the effective mass of an electron in the conduction band. This expression is based on the assumption that the density of energy levels in the conduction band is proportional to $\epsilon^{1/2}$ where ϵ is the energy of the level and that the probability of occupancy of the levels is determined by a Boltzmann distribution of the form $\exp(-\epsilon/\boldsymbol{k}T)$. In the present discussion, the zero of energy is taken as the bottom of the conduction band; the bound levels associated with the donors will be assumed at $-E_{\mathrm{D}}$ and the top of the valence band at $-E_{\mathrm{G}}$ is the forbidden bandwidth (figure 1).

Let there be N donor impurity atoms in the volume V of semiconductor and let i of these be occupied by electrons, i.e. neutral, the remaining $N-i$ being positive ions. The energy associated with the i electrons is $-iE_{\mathrm{D}}$ while the entropy

S is determined from the number of possible arrangements of the i electrons among the N levels taking into account the spin. Thus

$$F_{\mathrm{D}} = -\boldsymbol{k}T \ln \frac{N!}{i!\,(N-i)!} - \boldsymbol{k}T\sigma \ln 2 - iE_{\mathrm{D}} \qquad \ldots\ldots(3)$$

where σ is the spin term of the entropy and is equal to i for unpaired levels and to $N-i$ for paired levels. Using Stirling's approximation this becomes

$$F_{\mathrm{D}} = -\boldsymbol{k}T\left[N \ln \frac{N}{N-i} - i \ln \frac{i}{N-i} + \sigma \ln 2\right] - iE_{\mathrm{D}}. \qquad \ldots\ldots(4)$$

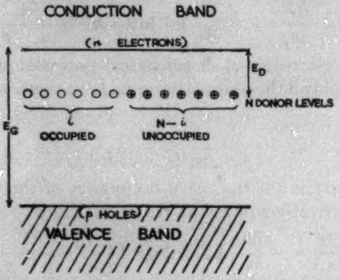


Figure 1. Schematic system of energy levels.

The free energy associated with the p missing electrons or holes in the valence band is derived in similar manner and is given by

$$F_{\mathrm{V}} = pE_{\mathrm{G}} - p\boldsymbol{k}T\left[\frac{3}{2} \ln \frac{2\pi m\,\boldsymbol{k}T}{\boldsymbol{h}^2} + 1 + \ln \frac{2V}{p}\right] \qquad \ldots\ldots(5)$$

where m_{p} is the effective mass of a hole in the valence band.

The constraint existing between the quantities n, i and p is

$$n + i = p + N \qquad \ldots\ldots(6)$$

and thus if fluctuations Δn in n and Δp in p occur there must simultaneously be a fluctuation $\Delta i = \Delta p - \Delta n$ in i.

Now the free energy may be expanded in terms of the first and second partial derivatives:

$$\begin{aligned} F = F_0 &+ \left[\frac{\partial F}{\partial n}\Delta n + \frac{\partial F}{\partial p}\Delta p + \frac{\partial F}{\partial i}\Delta i\right] \\ &+ \frac{1}{2}\left[\frac{\partial^2 F}{\partial n^2}(\Delta n)^2 + \frac{\partial^2 F}{\partial p^2}(\Delta p)^2 + \frac{\partial^2 F}{\partial i^2}(\Delta i)^2 + 2\frac{\partial^2 F}{\partial n \partial p}\Delta n\,\Delta p \right. \\ &\left. + 2\frac{\partial^2 F}{\partial n \partial i}\Delta n\,\Delta i + 2\frac{\partial^2 F}{\partial p \partial i}\Delta p\,\Delta i\right] \qquad \ldots\ldots(7) \end{aligned}$$

where F_0 is the value of the free energy when n, p and i have their most probable values. In thermal equilibrium the first order departure of F from F_0 must vanish whatever Δn, Δp and Δi whence from equations (6) and (7)

$$\frac{\partial F_{\mathrm{C}}}{\partial n} = -\frac{\partial F_{\mathrm{V}}}{\partial p} = \frac{\partial F_{\mathrm{D}}}{\partial i} \qquad \ldots\ldots(8)$$

which is the condition that the electronic electrochemical potentials μ must be the same for all three sets of levels. This leads to the equilibrium conditions:

$$\mu = -\boldsymbol{k}T\left[\frac{3}{2}\ln\frac{2\pi m_n\boldsymbol{k}T}{\boldsymbol{h}^2}+\ln\frac{2V}{n_0}\right] = -E_G+\boldsymbol{k}T\left[\frac{3}{2}\ln\frac{2\pi m_p\boldsymbol{k}T}{\boldsymbol{h}^2}+\ln\frac{2V}{p_0}\right]$$

$$= -E_D+\boldsymbol{k}T\ln\frac{i_0}{N-i_0}-\boldsymbol{k}T\frac{d\sigma}{di}\ln 2 \qquad \ldots\ldots(9)$$

where suffix 0 denotes the equilibrium value or mean value.

Adopting the customary notation:

$$N_C \equiv 2(2\pi m_n\boldsymbol{k}T/\boldsymbol{h}^2)^{3/2} \quad \text{and} \quad N_V \equiv 2(2\pi m_p\boldsymbol{k}T/\boldsymbol{h}^2)^{3/2} \qquad \ldots\ldots(10)$$

we have for the equilibrium conditions

$$\frac{n_0(N-i_0)}{i_0} = \beta V N_C \exp\left(-\frac{E_D}{\boldsymbol{k}T}\right) \qquad \ldots\ldots(11)$$

$$\frac{p_0 i_0}{N-i_0} = \frac{1}{\beta} V N_V \exp\left(-\frac{E_G-E_D}{\boldsymbol{k}T}\right) \qquad \ldots\ldots(12)$$

and

$$n_0 p_0 = V^2 N_C N_V \exp(-E_G/\boldsymbol{k}T) = n_i^2 \qquad \ldots\ldots(13)$$

where $\beta=2$ for paired electrons in the donor levels and $\frac{1}{2}$ for unpaired electrons. These equations are readily recognized as the mass action equations for the three types of transition:

conduction electron + positive donor ion $\longleftrightarrow$ neutral donor atom + E_D
hole + neutral donor atom $\longleftrightarrow$ positive donor ion + $E_G - E_D$
conduction electron + hole $\longleftrightarrow$ E_G.

The corresponding equations for the equilibrium numbers of electrons, holes and occupied (neutral) donor atoms are obtained directly from (9):

$$n_0 = V N_C \exp(\mu/\boldsymbol{k}T)$$

$$p_0 = V N_V \exp[-(E_G+\mu)/\boldsymbol{k}T]$$

$$i_0 = N[1+\beta\exp\{-(E_D+\mu)/\boldsymbol{k}T\}]^{-1}.$$

To evaluate the fluctuations we require the second derivatives of the free energy in the equilibrium condition which are given by

$$\frac{\partial^2 F_C}{\partial n^2} = \frac{\boldsymbol{k}T}{n_0}, \quad \frac{\partial^2 F_V}{\partial p^2} = \frac{\boldsymbol{k}T}{p_0}, \quad \frac{\partial^2 F_D}{\partial i^2} = \boldsymbol{k}T\left[\frac{1}{i_0}+\frac{1}{N-i_0}\right] \qquad \ldots\ldots(14)$$

whence the increase of F above F_0 due to fluctuations Δn and Δp is given by

$$\frac{F-F_0}{\boldsymbol{k}T} = \frac{(\Delta n)^2}{2n_0}+\frac{(\Delta p)^2}{2p_0}+\frac{(\Delta p-\Delta n)^2}{2}\left[\frac{1}{i_0}+\frac{1}{N-i_0}\right]. \qquad \ldots\ldots(15)$$

Now Δn and Δp are fluctuating variables with zero mean and have a bivariate normal distribution of the form $\exp(F_0-F)/\boldsymbol{k}T$ so that the above expression must be identified with

$$\frac{1}{1-r^2}\left[\frac{(\Delta n)^2}{2\sigma_n^2}+\frac{(\Delta p)^2}{2\sigma_p^2}-\frac{r\Delta n\Delta p}{\sigma_n\sigma_p}\right]$$

where σ_n and σ_p are the r.m.s. fluctuations in n and p, and r is the correlation coefficient between Δn and Δp defined by

$$r \equiv \overline{\Delta n\Delta p}/\sigma_n\sigma_p.$$

Thus the variances and covariances of Δn and Δp are given by

$$\overline{(\Delta n)^2}=\sigma_n{}^2 \quad =\frac{1}{1-r^2}\left[\frac{1}{n_0}+\frac{1}{i_0}+\frac{1}{N-i_0}\right]^{-1} \qquad \ldots\ldots(16)$$

$$\overline{(\Delta p)^2}=\sigma_p{}^2 \quad =\frac{1}{1-r^2}\left[\frac{1}{p_0}+\frac{1}{i_0}+\frac{1}{N-i_0}\right]^{-1} \qquad \ldots\ldots(17)$$

$$\overline{\Delta n\Delta p}=r\sigma_n\sigma_p=\frac{r^2}{1-r^2}\left[\frac{1}{i_0}+\frac{1}{N-i_0}\right]^{-1} \qquad \ldots\ldots(18)$$

where

$$r^2=\left[\frac{1}{i_0}+\frac{1}{N-i_0}\right]^2\left[\frac{1}{n_0}+\frac{1}{i_0}+\frac{1}{N-i_0}\right]^{-1}\left[\frac{1}{p_0}+\frac{1}{i_0}+\frac{1}{N-i_0}\right]^{-1} \qquad \ldots.(19)$$

and

$$\overline{(\Delta i)^2}=\overline{(\Delta p-\Delta n)^2}=\sigma_n{}^2+\sigma_p{}^2-2r\sigma_n\sigma_p. \qquad \ldots..(20)$$

These equations completely specify the fluctuations in the number of electrons in the conduction band, and in the number of holes in the valence band and also the correlation between these fluctuations. The mean deviation of the free energy above its minimum value is given by $\overline{F-F_0}=\boldsymbol{k}T$ which is the value appropriate to a system whose state is determined by two variables, i.e. any two of n, p and i.

Some inferences will now be drawn from these general relations:

(i) In *intrinsic* material we are only concerned with Δn and Δp whose correlation coefficient is unity since $\Delta n=\Delta p$ at all times. Furthermore

$$\overline{(\Delta n)^2}=\overline{(\Delta p)^2}=\left[\frac{1}{n_0}+\frac{1}{p_0}\right]^{-1}=\tfrac{1}{2}n_0=\tfrac{1}{2}p_0=\tfrac{1}{2}n_i \qquad \ldots\ldots(21)$$

so that the mean square fluctuation in n and p is equal to half their equilibrium value.

(ii) In strongly *extrinsic n-type* material the number of holes p_0 will be very much smaller than the number of electrons n_0 and we have three cases:

(*a*) Low ionization of donor atoms, $p_0 \ll N-i_0=n_0-p_0 \simeq n_0 \ll i_0$ whence $r^2 \simeq p_0/2n_0$

$$\overline{(\Delta i)^2}\simeq\overline{(\Delta n)^2}\simeq\tfrac{1}{2}n_0 \qquad \ldots\ldots(22)$$

$$\overline{(\Delta p)^2}\simeq p_0. \qquad \ldots\ldots(23)$$

Thus the mean square fluctuation in the number of conduction electrons is approximately one half the mean number.

(*b*) High ionization of donor atoms: $p_0 \ll i_0 \ll N-i_0=n_0-p_0$. Then

$$r^2 \simeq p_0/i_0 \ll 1$$

$$\overline{(\Delta i)^2}\simeq\overline{(\Delta n)^2}\simeq i_0 \simeq N-n_0 \qquad \ldots\ldots(24)$$

$$\overline{(\Delta p)^2}\simeq p_0. \qquad \ldots\ldots(25)$$

Thus the mean square fluctuations in the number of electrons in the conduction band is approximately equal to the number of un-ionized donor atoms.

(*c*) In the case of partial ionization of the donor atoms which lies intermediately between and includes the two limiting cases (*a*) and (*b*) we have

$$\overline{(\Delta n)^2}\simeq\left[\frac{2}{n_0}+\frac{1}{N-n_0}\right]^{-1} \qquad \ldots..(26)$$

so long as p_0 is negligible. Figure 2 shows $\overline{(\Delta n)^2}/n_0$ as a function of n_0/N.

(iii) The case of the *nearly intrinsic semiconductor* is of considerable practical interest because much of the experimental work on fluctuation noise in germanium filaments has been done on single-crystal material of high purity. However it is also the case in which it is most difficult to approximate to the exact equations (16) to (20), without loss of accuracy since n_0, p_0 and $N-i_0$ are all comparable in magnitude. If it be assumed that nearly all the donors are ionized, which is quite reasonable for pure germanium at room temperature, we can regard i_0 as small compared with N, n_0 and p_0 and furthermore $n_0-p_0=N$. With this assumption we find that

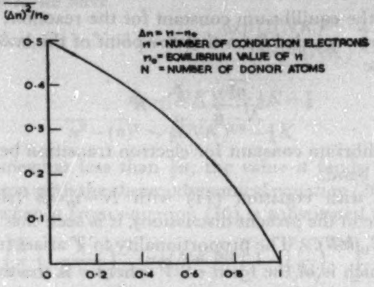


Figure 2. Mean square fluctuation in the number of conduction electrons in an n-type extrinsic semiconductor with a negligible number of holes.

$$\overline{(\Delta n)^2}=\overline{(\Delta p)^2}=n_0p_0/(n_0+p_0) \qquad \ldots\ldots(27)$$

and

$$\overline{(\Delta i)^2}=i_0. \qquad \ldots\ldots(28)$$

The fluctuation in n has the expected behaviour for perfectly intrinsic material, giving $\overline{(\Delta n)^2}=\frac{1}{2}n_0$ (compare equation (21)) and for appreciably extrinsic material giving $\overline{(\Delta n)^2}=n_0$ or p_0 whichever is the smaller. Thus so long as the impurity atoms are nearly completely ionized so that *the number of neutral atoms is small compared with the number of minority carriers*, the mean square fluctuation of n and p simply tends to the number of minority carriers.

§ 3. STATISTICAL TREATMENT

To consider a statistical model which can be applied to the thermal ionization of extrinsic n-type semiconductors, let N be the number of donor atoms in a volume V and let at any given moment n be the number of electrons raised from the donor atoms to the conduction band; then n will also be the number of positive donor ions and $i=N-n$ the number of neutral donor atoms since it is now assumed that the number of holes is negligible. The probability that one of the neutral atoms will be ionized in a time interval dt is

$$g(n)\,dt=\gamma(N-n)\,dt \qquad \ldots\ldots(29)$$

while the probability that one of the n conduction electrons will recombine with one of the n positive ions during the same interval dt is

$$r(n)\,dt=\rho n^2\,dt \qquad \ldots\ldots(30)$$

where γ and ρ are constants independent of n. In thermal equilibrium the processes of ionization and recombination must balance statistically:

$$\frac{d\bar{n}}{dt} = g_0 - r_0 = \gamma(N-\bar{n}) - \rho\overline{n^2} = 0.$$

The ratio γ/ρ is the equilibrium constant for the reaction: electron + positive ion $\longleftrightarrow$ neutral atom, regarded from the viewpoint of the law of mass action and we accordingly write

$$\frac{\overline{n^2}}{N-\bar{n}} = \frac{\gamma}{\sigma} \equiv K \qquad \text{.......(31)}$$

where K is the equilibrium constant for electron transition between donor atoms and the conduction band.

By comparison with equation (11) with $N-i_0=\bar{n}$ (since the holes are considered negligible in the present discussion), it is seen that K can be identified with $\beta V N_C \exp(-E_D/\boldsymbol{k}T)$. The proportionality to V arises from the recombination parameter ρ which is of the form $\overline{vS}/V$ where v is the thermal velocity of a conduction electron, and S is its cross section for capture by a positive ion, the mean being taken over all values of v.

From the dependence of g and r on n the equilibrium probability distribution function $P(n)$ can be determined from the requirement that it shall be statistically stationary. Since $\overline{n^2}=K(N-\bar{n})$ it is evident that the following inequalities hold:

$$(\overline{n^2})^{1/2} > \{(\tfrac{1}{4}K^2+KN)^{1/2}-\tfrac{1}{2}K\} > \bar{n}.$$

Now it follows from the difference equation

$$g(n-1)P(n-1)+r(n+1)P(n+1)=[g(n)+r(n)]P(n)$$

that in the steady state

$$P(n)=P(0)\frac{\prod\limits_{\nu=0}^{n-1} g(\nu)}{\prod\limits_{\nu=1}^{n} r(\nu)} = \frac{N!\,K^n}{(N-n)!\,n!^2}P(0).$$

From $\sum\limits_{n=0}^{N} P(n)=1$ it is seen that $\dfrac{1}{P(0)} = \sum\limits_{n=0}^{N} \dfrac{N!\,K^n}{(N-n)!\,n!^2} \equiv M$

whence

$$P(n)=\frac{1}{M}\frac{N!\,K^n}{(N-n)!\,n!^2} \qquad \text{......(32)}$$

where M denotes the confluent hypergeometric function, ${}_1F_1(-N;1;-K)$. Thus

$$\bar{n}=\sum_{n=1}^{N} nP(n)=\frac{K}{M}\frac{dM}{dK} \qquad \text{......(33)}$$

$$\overline{n^2}=\sum_{n=1}^{N} n^2P(n)=\frac{1}{M}\left(K^2\frac{d^2M}{dK^2}+K\frac{dM}{dK}\right). \qquad \text{......(34)}$$

Combining these two latter equations with (31) is consistent with the second order differential equation for M, the prime denoting differentiation with respect to K:

$$KM''+(K+1)M'-NM=0.$$

From equations (33) and (34) we have quite generally for the mean square fluctuation:

$$\overline{n^2}-(\bar{n})^2=K^2\frac{d^2}{dK^2}(\ln M)+K\frac{d}{dK}(\ln M). \qquad \ldots\ldots(35)$$

(i) Now for $N \gg K$ we have

$$M \equiv {}_1F_1(-N;1;-K) \sim \frac{\exp[2(NK)^{1/2}-\frac{1}{2}K]}{(4\pi)^{1/2}(NK)^{1/4}}$$

whence

$$\bar{n} \sim (NK)^{1/2}-\tfrac{1}{2}K-\tfrac{1}{4} \qquad \ldots\ldots(36)$$

$$\overline{n^2}-(\bar{n})^2 \sim \tfrac{1}{2}(NK)^{1/2}-\tfrac{1}{2}K \qquad \ldots\ldots(37)$$

which is always somewhat less than $\frac{1}{2}\bar{n}$, the value it tends to as N/K tends to infinity. Comparison with the thermodynamical equation (26) for the fluctuation shows agreement, when $\bar{n}$ from equation (36) is substituted for n_0:

$$\left(\frac{2}{\bar{n}}+\frac{1}{N-\bar{n}}\right)^{-1} \simeq \tfrac{1}{2}(NK)^{1/2}\left[1-\left(\frac{K}{N}\right)^{1/2}\right].$$

Although these results for weak ionization in an extrinsic semiconductor have been derived for the case of a negligible number of holes it is readily seen that the argument may be applied to an intrinsic semiconductor, if N is identified with the number of semiconductor atoms which will greatly exceed the number of conduction electrons n which will in this case be equal to the number of holes, p. This similarity arises because the fundamental generation and recombination equations (29) and (30) are applicable, with the generation rate g effectively constant since $N/\bar{n}$ would usually be at least 10^3.

(ii) When the degree of donor ionization in an extrinsic semiconductor is large ($K \gg N$), the mean number of conduction electrons approaches that of the donor atoms and the appropriate asymptotic development of M is that using the final terms in the series development

$${}_1F_1(-N;1;-K) \sim \frac{K^N}{N!}\left[1+\frac{N^2}{K}+\frac{N^2(N-1)^2}{2K^2}+\frac{N^2(N-1)^2(N-2)^2}{6K^3}+\ldots\right]$$

whence

$$\bar{n}=N-\frac{N^2}{K}+\frac{N^2(2N-1)}{K^2}+\frac{N^2(5N^2-6N+2)}{K^3}+\ldots \qquad \ldots\ldots(38)$$

$$\overline{n^2}=N^2-\frac{N^2(2N-1)}{K}+\frac{N^2(5N^2-6N+2)}{K^3}-\ldots \qquad \ldots\ldots(39)$$

and thus

$$\overline{n^2}-(\bar{n})^2=\frac{N^2}{K}\left[1-\frac{2(2N-1)}{K}+\ldots\right]. \qquad \ldots\ldots(40)$$

Hence as K/N tends to infinity

$$N-\bar{n} \to N^2/K \quad \text{and} \quad \overline{n^2}-(\bar{n})^2 \to N^2/K$$

which implies that

$$\overline{i^2}-(\bar{i})^2=\overline{(N-n)^2}-(N-\bar{n})^2 \to N-\bar{n}=\bar{i} \qquad \ldots\ldots(41)$$

and the number of un-ionized donor atoms $N-n=i$ has the attribute of a Poisson distribution with a mean value N^2/K, since its variance is equal to its mean value.

Again we may compare the result obtained statistically (equation (40)) with that derived thermodynamically (equation (26)):

$$\left[\frac{2}{n_0}+\frac{1}{N-n_0}\right]^{-1} \simeq \left[2\left\{N-\frac{N^2}{K}\right\}^{-1}+\left\{\frac{N^2}{K}-\frac{N^2(2N-1)}{K^2}\right\}^{-1}\right]^{-1} \simeq \frac{N^2}{K}\left[1-\frac{4N}{K}\right]$$

which agrees with (40) to the first order correction term in N/K.

It is seen that the merit of the thermodynamical approach is that the general case may be treated of an extrinsic semiconductor in which N/K is neither very small nor very large compared with unity yielding the comprehensive formula (26). Furthermore the presence of a significant number of holes in the valence band can also be readily yielded in the thermodynamical approach as was shown in equations (16)–(20), and both the variances and covariances of the numbers of electrons in the various bands may be derived.

(iii) Lastly we consider the semiconductor in which all the N donor atoms are ionized so that the numbers of conduction electrons and holes in the valence band are related by

$$n = N + p. \qquad \ldots\ldots(42)$$

The rate of hole–electron pair generation will be assumed to be independent of n:

$$g = \text{constant} \qquad \ldots\ldots(43)$$

while the recombination coefficient will be assumed proportional to the product of p and n:

$$r = \rho p n = \rho p(N+p). \qquad \ldots\ldots(44)$$

In thermodynamical equilibrium we see that g and ρ must be related by

$$\frac{g}{\rho} = \overline{pn} = n_i^2 \qquad \ldots\ldots(45)$$

where n_1 is the r.m.s. number of electrons (or holes) which would exist in the semiconductor (of volume V) if it were intrinsic (cf. equation (13)).

The probability distribution function for the number of holes is found from

$$P(p) = \frac{N!\,(g/\rho)^p}{p!\,(N+p)!}\,P(0)$$

whence

$$P(p) = \frac{n_1^{N+2p}}{p!\,(N+p)!\,\mathrm{I}_N(2n_1)} \qquad \ldots\ldots(46)$$

where I_N is the modified Bessel function. Thus

$$\bar{p} = n_1\frac{\mathrm{I}_{N+1}(2n_1)}{\mathrm{I}_N(2n_1)} \qquad \ldots\ldots(47)$$

and

$$\overline{p^2} = \frac{n_1^2\mathrm{I}_{N+2}(2n_1)+n_1\mathrm{I}_{N+1}(2n_1)}{\mathrm{I}_N(2n_1)} \qquad \ldots\ldots(48)$$

are found from $P(p)$. It is readily shown that the equilibrium relation $n_1^2 = N\bar{p}+\overline{p^2}$ from equations (43), (44) and (45) agrees with the recurrence formulae for Bessel functions.

In general the mean square fluctuation in the number of holes or electrons is

$$\overline{n^2}-(\bar{n})^2 = \overline{p^2}-(\bar{p})^2 = n_1^2 - N\bar{p}-(\bar{p})^2 = n_1^2 - Nn_1\frac{\mathrm{I}_{N+1}}{\mathrm{I}_N} - n_1^2\frac{\mathrm{I}_{N+1}^{\,2}}{\mathrm{I}_N^{\,2}}.$$

In order to approximate satisfactorily to this expression it is relevant to know the relative magnitude of N and $2n_1$ that is to say whether the semiconductor is very nearly intrinsic ($2n_1 \gg N$) or largely extrinsic ($2n_1 \ll N$) since these conditions influence the choice of asymptotic expansion used for the Bessel function.

If the material is nearly intrinsic we find that the mean values are given by

$$\bar{n} - N = \bar{p} = n_1 - \tfrac{1}{2}N - \tfrac{1}{4} + (4N^2 - 1)/32n \qquad \ldots\ldots(49)$$

whence

$$n^2 - (n)^2 = p^2 - (\bar{p})^2 = \tfrac{1}{2}n_1 \qquad \ldots\ldots(50)$$

in accordance with earlier results (equation (21)) for a perfectly intrinsic semiconductor. The slow dependence of the fluctuation on the presence of a small number of donor atoms arises from the fact that the fluctuation has its maximum in the intrinsic condition (cf. equation (27)).

In a paper published elsewhere (Burgess 1954) an asymptotic method of applying the statistical approach is developed which leads to very simple general formulae for the variance of the number of electrons and for the relaxation time of a system when slightly perturbed from equilibrium.

ACKNOWLEDGMENTS

The work described above was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

BURGESS, R. E., 1954, *Physica*, **20**, 1007.
MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford: Clarendon Press), Chap. V.
VAN DER ZIEL, A., 1953 a, *J. Appl. Phys.*, **24**, 222; 1953 b, *Ibid.*, **24**, 1063.

A Theory of Action-at-a-Distance

By G. BURNISTON BROWN
University College, University of London

MS. received 17th March 1955

Abstract. A theory of non-instantaneous action-at-a-distance and physical relativity is used to develop a force-formula for particles in relative motion. The velocity terms are derived from the forces between electrical circuits, and the acceleration terms from experiments with oscillators. The formula so obtained, when corrected for 'retardation' accounts for electromagnetic induction.

On the hypothesis that the variation of force with motion is the same for gravitational forces, an explanation of inertia is given in terms of the total amount of matter in the universe and its distribution, in quantitative agreement with present estimates of the mean density. A 'red-shift' near large bodies and a perihelion motion of the planets also follows. This unified theory of gravitational and electrical force appears capable of giving a physical explanation of the macroscopic phenomena of physics.

§ 1. INTRODUCTION

It is well known that the idea of action-at-a-distance has usually met with repugnance, and that Newton rejected it. Physicists have preferred emission theories or medium theories, it being generally assumed that 'contact' forces do not require further elucidation. It was never clear, however, that contact was not action-at-a-distance with the distance very small.

An examination of the application of the hypothesis of action-at-a-distance to macroscopic physics is the object of the present paper; it is defined as follows:

Action-at-a-distance.

The forces exerted between particles of matter act at a distance, i.e. without the necessity for any intervening matter.

To this is added

Non-instantaneous action.

All particles in the universe are continually interacting, but a particle at time t can only affect another particle with respect to which it is in motion at a distance r, at a time $(t+r/c)$ later (where $c \simeq 3 \times 10^{10}$ cm sec^{-1}).

Newton, following Galileo, treated forces as acting independently, so that the force between two particles depended only on the particles themselves and their relative distance apart, and this principle of independence and relativity is extended to include motion as follows:

The principle of physical relativity.†

Every particle acts on every other particle with a force which depends only on the particles themselves, their relative separation and motion, and the constant of interaction c.‡

---

† So called to distinguish it from the principle of relativity which is concerned with the relations between moving observers' measurements.

‡ For the advantages of referring to c as the constant of interaction rather than as the velocity of anything, see Brown (1941).

These, together with the principle of superposition, are the postulates of this theory of non-instantaneous action-at-a-distance and physical relativity. We now attempt to determine, from information derived from experiment and observation, the extra terms required to extend the macroscopic force-formula to include the effects produced by relative motion, just as Newton, in finding the first term, used the results of Kepler and Tycho Brahe.

§ 2. VELOCITY TERMS

We start with the fact that when two particles are at rest relatively to one another, the macroscopic interactions, gravitational and electrical, are represented by similar force-formulae viz. mm'/r^2 and qq'/r^2 where m, m' are the masses in dynamical units, q, q' are the charges in e.s.u., and r is the distance between them. This similarity suggests the hypothesis that the variation of force with motion is also similar in the two cases.

Proceeding, then, on this hypothesis, we further assume, with Ampère, that all magnetic effects are the result of the motion of charges, and that in neutral conductors carrying steady currents, we have charged particles moving with an effective resultant 'drift' velocity v, and in oscillators we have charges moving with acceleration f as well.

Commencing with the v terms, we take "the most general form of Ampère's law consistent with the experimental facts" (Maxwell 1904). For steady, uniform currents when one, at least, of the circuits is closed, we have for the force in the x direction between two elements ds, ds' carrying currents i and i', distant r apart:

$$F_x = -ii'\,ds\,ds'[R\cos(rx) + S\cos(xds) + S'\cos(xds')]$$

where

$$R = \frac{1}{r^2}\left(\frac{\partial r}{\partial s}\frac{\partial r}{\partial s'} - 2r\frac{\partial^2 r}{\partial s\,\partial s'}\right) + r\frac{\partial^2 Q}{\partial s\,\partial s'}, \qquad S = \frac{\partial Q}{\partial s'}, \qquad S' = -\frac{\partial Q}{\partial s}.$$

The added Q terms give zero when integrated round a complete circuit, and are therefore compatible with Ampère's results. Taking $Q = K/r$ where K is some constant, we can show that

$$F_x = -\frac{ii'\,ds\,ds'}{r^2}\{[(K+2)\cos(dsds') - 3(K+1)\cos(rds)\cos(rds')]\cos(rx) + K\cos(rds')\cos(xds) + K\cos(rds)\cos(xds')\}.$$

To convert this expression for F_x into one involving charges q, q' and velocities v and v', we make use of the experimentally determined relation $qv/c = ids$ (Rowland and Hutchinson 1889). Thus for the force we write

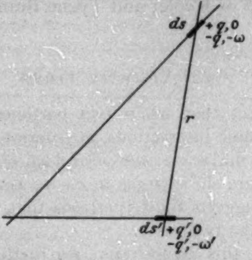
$$E_x = \frac{-qq'}{r^2c^2}\{[(K+2)vv'\cos(dsds') - 3(K+1)v_rv_r']\cos(rx) + Kv_x'v_r + Kv_xv_r'\} \qquad \ldots\ldots(1)$$

since v is along ds and v' along ds'.

With physical relativity, only the relative velocity affects the force, and so, noting that dimensionally the terms are [velocity²], we shall have terms in v^2, v_r^2, and v_xv_r. To find the coefficients of these terms we must calculate the resultant force of one neutral element of current on another neutral element.

§ 3. FORCE BETWEEN LINEAR CIRCUITS AT REST

We now have to consider the neutral current elements *ids* and *i′ds′*, as consisting of positive charges $+q$, $+q'$ at rest, and negative charges $-q$, $-q'$ in motion (see figure). Let the velocity of the negative charge in *ds* be $-\omega$ and in *ds′*, $-\omega'$.



Taking axes in the conductor element *ds*, the force of the element *ds′* on *ds* will be the resultant of four forces: the forces on $+q$ and $-q$ (velocity $-\omega$) at *ds*, due to $+q'$ (velocity zero) and $-q'$ (velocity $-\omega'$) at *ds′*.
The terms in v^2 are

$$\begin{aligned}\sum qq'v^2 &= -qq'(\omega_x')^2 - qq'(\omega_x)^2 + qq'(\omega_x - \omega_x')^2 + \dots + \dots \\ &= -2qq'\sum \omega_x\omega_x' = -2c^2 ii'\,ds\,ds' \cos(ds\,ds').\end{aligned}$$

Similarly

$$\sum qq'v_r^2 = -2c^2 ii'\,ds\,ds' \cos(r\,ds)\cos(r\,ds')$$

$$\sum qq'v_xv_r = -c^2 ii'\,ds\,ds'[\cos(x\,ds)\cos(r\,ds') + \cos(r\,ds)\cos(x\,ds')].$$

Comparing these terms with the most general expression of Ampère's formula (1), we see that if we multiply the v^2 terms by $-(K+2)/2c^2$, the v_r^2 terms by $-3(K+1)/2c^2$, and the v_xv_r terms by $-K/c^2$, we shall obtain agreement. Thus we have for the velocity terms:

$$E_x = \frac{qq'}{r^2}\left\{\left[\frac{(K+2)}{2}\frac{v^2}{c^2} - \frac{3(K+1)}{2}\frac{v_r^2}{c^2}\right]\cos(rx) + \frac{Kv_xv_r}{c^2}\right\}. \qquad \dots\dots(2)$$

§ 4. ACCELERATION TERMS

In order, next, to derive the acceleration terms, the dipole oscillator suggests itself as perhaps the simplest case to take; and if we consider the interaction effects produced at distances large compared with the amplitude of the vibration of the current, then using polars, we can treat r and θ as the same for all parts of the oscillator, and the only effect of 'retardation' (i.e. the finite interaction constant c) will be in the phase.

We shall assume that experiment shows (Hertz 1900, Ratcliffe 1931–2, McPetrie, Perry and Ford 1945) that a dipole oscillator, at large distances, only produces electric force at right angles to the radius vector r in a plane containing r and the direction of oscillation, of amount $E = (qap^2 \sin\theta / rc^2)\cos p(t - r/c)$ where a is the amplitude of the oscillation of a charge q, p is the pulsatance, and θ the angle between r and the direction of oscillation. Neglecting the phase retardation, and remembering that f (the acceleration) is $-ap^2 \cos pt$, we find that $E_x = (qq'/rc^2)[f_r \cos(rx) - f_x]$ satisfies the conditions.

An oscillating dipole is not, of course, a neutral conductor: there are free fluctuating charges at its ends, but the force due to these decreases as $1/r^3$ and so rapidly becomes negligible compared with $1/r$. The velocity terms compared with the acceleration terms are of the order v^2/rf or $a^2p^2/ap^2r = a/r$ and are therefore negligible.

Having obtained the acceleration terms from experiments in which r and $\cos(rx)$ are sensibly the same for all parts of the circuit, so that retardation effects due to changes in them are not appreciable, we must now consider the case in which this is not so, e.g. ordinary induction with linear circuits in the laboratory.

Now retardation makes the effect at time t of a particle P on a particle Q (with respect to which it is moving), depend on the position and motion of P at a time $(t - R/c)$ where R was the distance separating P and Q at that time. We can use Taylor's theorem in the form

$$x\left(t - \frac{R}{c}\right) = x(t) - \frac{R}{c} v_x(t) + \frac{R^2}{2c^2} f_x(t) + \dots$$

$$v_x\left(t - \frac{R}{c}\right) = v_x(t) - \frac{R}{c} f_x(t) + \dots$$

to express the 'retarded' function $\cos(Rx)/R$ in terms of the instantaneous values $\cos(rx)/r$, provided the expansions are rapidly convergent. Assuming that this condition is satisfied in practice, we can show that for circuits in which r and $\cos(rx)$ vary, and therefore cause varying retardation, the acceleration terms are

$$-\frac{qq'}{2rc^2}[f_r \cos(rx) + f_x]. \qquad \dots\dots(3)$$

§ 5. ELECTROMAGNETIC INDUCTION

When we consider the inductive effect in an element of a circuit due to another element, in the same or another circuit, we have to find the force tending to separate the positive and negative charges in the element in the direction of the inducing element, and sum this up for the whole circuit. From this force, when it refers to unit charge, can be found the induced electromotive force.

It is not difficult to show that the velocity terms and the acceleration terms derived above, both, independently, yield e.m.f. $= -dN/dt$ for the induced electromotive force.

§ 6. THE LAW OF GRAVITATION AND THE MOTION OF MERCURY

We have now obtained from electrodynamic experiments the extra terms involving the velocity (equation (2)) and the acceleration (equation (3)) which have to be added to the electrostatic force-formula in order to account for the facts of electromagnetism. If we proceed by making the hypothesis, already mentioned, that gravitational force varies in the same way, we can obtain the gravitational force between two masses m and m' (dynamical units) by merely substituting m and m' for e and e', and changing the sign, and we then have

$$F_x = -\frac{mm'}{r^2}\left\{\left[1 + \frac{(K+2)}{2}\frac{v^2}{c^2} - \frac{3(K+1)}{2}\frac{v_r^2}{c^2} - \frac{rf_r}{2c^2}\right]\cos(rx) + \frac{Kv_x v_r}{c^2} - \frac{rf_x}{2c^2}\right\}.$$

Now, if a case were known where there was a discrepancy between the Newtonian formula and observation, and this discrepancy could be measured with a fair degree of accuracy, we might be able to determine the value of the constant K. The movement of the perihelion of Mercury provides such an instance. In this planetary case the acceleration is negligible except along the radius vector, and this is given very closely by $-GM/r^2$ where M = mass of the Sun in grammes. If this value is substituted in the acceleration terms, the force-formula leads to a rotation of the perihelion per revolution of

$$\frac{(3-K)\pi GM}{c^2 a(1-e^2)}$$

where e is the eccentricity and a is the semi-major axis of the orbit. If we take $K=-3$ we get the formula first obtained by Gerber in 1898, and by Einstein in 1916. Ritz obtained a similar formula in 1908 but neglected the acceleration terms. We are not, of course, compelled to choose K a whole number, but the value -3 gives agreement with experiment within the limits of accuracy:

Observed value for Mercury	$42''{\cdot}56 \pm 0{\cdot}94$ per century (Clemence 1947)
Calculated ($K=-3$)	$43''{\cdot}03 \pm 0{\cdot}03$

Thus the force-formulae become

$$E_x = \frac{qq'}{r^2}\left\{\left[1-\frac{1}{2}\frac{v^2}{c^2}+\frac{3v_r^{\,2}}{c^2}-\frac{rf_r}{2c^2}\right]\cos(rx)-\frac{3v_x v_r}{c^2}-\frac{rf_x}{2c^2}\right\} \qquad \ldots\ldots(4)$$

$$F_x = -\frac{mm'}{r^2}\left\{\left[1-\frac{1}{2}\frac{v^2}{c^2}+\frac{3v_r^{\,2}}{c^2}-\frac{rf_r}{2c^2}\right]\cos(rx)-\frac{3v_x v_r}{c^2}-\frac{rf_x}{2c^2}\right\}. \qquad \ldots\ldots(5)$$

§ 7. ORIGIN OF INERTIAL FORCES

Having obtained a formula for the interaction of uncharged particles of matter which takes account of relative velocity and acceleration, we can inquire into the problem of *inertia*. If this is not due to movement with respect to 'absolute space', it ought to be due to surrounding matter, as suggested by Bishop Berkeley when criticizing Newton, and later by Mach. Now the evidence of astronomical observation at the present time is that the matter of the universe is distributed more or less uniformly, and to about the same distance in all directions. We must therefore consider the force on a moving body at the centre of a spherical distribution of matter of uniform density ρ (dynamical units) and radius R. Using the postulate of physical relativity, we can take our particle of mass m to be at the centre of coordinates, and the universe moving in the opposite direction. On calculating the force by equation (5) we find that for a steady velocity the force of the universe on m is zero, but for an acceleration f there is an opposing force equal to $-(4/3)(\pi m\rho R^2/c^2)\times f$. If we take this to be the force of inertia and write m_1 for the inertial mass, we shall have

$$F = m_1 f = \frac{4}{3}\frac{\pi\rho R^2}{c^2}mf.$$

Thus the ratio of the attractive mass to the inertial mass of a body which we know to be $\sqrt{G}$ should be given by $3c^2/4\pi\rho R^2$ or

$$G = \frac{9c^4}{16\pi^2\rho^2 R^4}. \qquad \ldots\ldots(6)$$

Taking $G=6{\cdot}7\times10^{-8}$ and $R=2\times10^{27}$ cm we can calculate the mean density of matter in the universe from equation (6) which yields 10^{-27} g cm^{-3}, a result which agrees with present estimates (Zwicky 1952).

§ 8. MASS AND RADIUS OF A SPHERICAL UNIVERSE AND RED-SHIFT

With ρ in usual units, equation (6) becomes $GM/c^2=R$, where M = mass of the universe. For the 'Einstein universe' $GM/c^2=1{\cdot}57R$, (Eddington 1949) and Whitrow (1950), as a result of a "concise argument partly based on Newtonian mechanics" gets $GM/c^2=1{\cdot}67R$.

Equation (6) also allows the increased inertial effect near a spherical body of mass m_s to be calculated, and the effect of this on the frequency of a rotating or oscillating particle can be shown to be a 'red-shift' by a factor $1-\frac{1}{2}\rho_s r_s^2/\rho R^2$ where ρ_s and r_s are the density and radius of the body. In the case of the Sun this is of the order of one part in a million. As is known, a shift of this order is present on the Sun's limb.

§ 9. THE FORCE ON A MOVING ELECTRON

The velocity terms in the force-formula have been derived from experiments on the force between current circuits, where the velocities of the moving charges concerned are, as far as we know, small compared with c. It is interesting to see how far the formula holds for moving free electrons. If we take, first, the case where the electron moves with velocity v at right angles to an electric field, as it does in experiments of the Bucherer type, we can show that the force on it is no longer $4\pi\sigma e$ (where the symbols have the usual meanings) but $4\pi\sigma e(1+\frac{1}{2}v^2/c^2)$.

Dividing by m_1, the acceleration could be written (if v^2/c^2 is small)

$$f=\frac{4\pi\sigma e}{m_1(1-v^2/c^2)^{1/2}}$$

so that if the variation of electric force with velocity is treated as a change in m_1, this would agree with $m=m_0(1-v^2/c^2)^{-1/2}$. Unfortunately the accuracy of this type of experiment is not sufficient (cf. Zahn and Spees 1938) to decide whether the force-formula is correct, or whether it requires the addition of higher powers of v^2/c^2. The case where the electron moves in the direction of the field, as in accelerators, is difficult to deal with, owing to lack of information on the rapidly varying non-uniform fields employed, but the formula indicates a limiting velocity. The force of a neutral current (or magnet) on a moving charged particle is correctly given by the formula.

§ 10. OPTICAL EFFECTS PRODUCED BY MOTION

Action-at-a-distance avoids the need for an ether, and therefore for 'waves'. Particles interact directly with one another, but the presence of other particles in the neighbourhood may, by superposition, cause a change of phase, which we are accustomed to look upon as a 'change in the velocity of light' (cf. Jenkins and White 1951) although, in all interaction, there is no change in the constant c. Bearing this in mind it is not difficult to show that the Doppler effect, aberration, and phenomena in moving bodies, can be adequately explained in terms of action-at-a-distance.

The Michelson–Morley and Kennedy–Thorndike experiments yield results which would be expected, since without an ether, movement is only with respect to the celestial bodies, and uniform motion with respect to these produces, as we have seen, no forces, and consequently no changes. The experiments of Ives and Stilwell, Michelson and Gale, Sagnac, and Dufour and Prunier, require a more detailed criticism which it is hoped to publish at a later date, when it will be shown that they do not invalidate the present theory.

§ 11. CONCLUSION

Experiment shows that electrical oscillations produce a very small force in a radial direction (Cullen 1952) (radiation pressure) so that the force-formula requires a small extra term, and phenomena occurring when $v \to c$ may also need extra terms. When more accurate experimental results become available it will be possible to decide what these terms should be. The theory of non-instantaneous action-at-a-distance and physical relativity seems capable of giving a causal explanation of the phenomena of macroscopic physics in terms of matter, force and motion; and (with the above reservations) the formula found gives quantitative agreement with experiment.

ACKNOWLEDGMENTS

The author's best thanks are due to Professor H. T. Flint for several helpful discussions, to Dr. R. Capildeo for checking some of the mathematics, and to Dr. O'Rahilly for his stimulating volume *Electromagnetics*.

REFERENCES

BROWN, G. B., 1941, *Proc. Phys. Soc.*, **53,** 418.
CLEMENCE, G. M., 1947, *Rev. Mod. Phys.*, **19**, 361.
CULLEN, A. L., 1952, *Proc. Instn Elect. Engrs*, Pt. IV, **99,** Monograph No. 23.
EDDINGTON, Sir A. S., 1949, *Fundamental Theory* (Cambridge : University Press), p. 10.
HERTZ, H., 1900, *Electric Waves* (London : Macmillan).
JENKINS, F. A., and WHITE, H. E., 1951, *Fundamentals of Optics*, 2nd Edn (London : McGraw-Hill), p. 459.
MAXWELL, J. C., 1904, *Treatise on Electricity and Magnetism*, II, 173.
MCPETRIE, J. S., PERRY, W. E., and FORD, L. E., 1945, *Wireless Engr*, **22,** 6.
RATCLIFFE, J. A., 1931–2, *J. Instn Elect. Engrs*, **70,** 528.
ROWLAND, H. A., and HUTCHINSON, C. T., 1889, *Phil. Mag.*, **27,** 445.
WHITROW, G. J., 1950, *The Structure of the Universe* (London : Hutchinson), p. 101
ZAHN, C. T., and SPEES, A. H., 1938, *Phys. Rev.*, **53,** 511.
ZWICKY, F., 1952, *Publ. Astr. Soc. Pacif.*, **64,** 242.

RESEARCH NOTES

The Field Distribution in Asymmetrical Magnetic Electron Lenses

By G. LIEBMANN
Research Laboratory, Associated Electrical Industries Limited, Aldermaston, Berks.

MS. received 1st *March* 1955

Symmetrical magnetic electron lenses, in which both pole pieces have the same bore diameter, have been extensively studied, and their properties are now well known. In practice, it is often necessary to deviate considerably from the symmetrical lens design; for instance, an asymmetrical form of lens, as shown in figure 1, is often used for electron microscope objective lenses where the enlarged half of the lens accommodates the specimen holder.

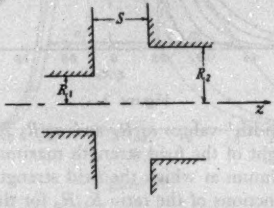


Figure 1.

The only information published on the field distribution in asymmetrical magnetic lenses seems to be the measurement taken by Dosse (1941) on a particular magnetic lens, and the calculation by Lenz (1950) for a very similar one.

To obtain a firm basis for the further extension of our knowledge of magnetic electron lenses, a systematic investigation was made of the field distribution in asymmetrical electron lenses and particularly of its dependence on the degree of asymmetry. The method used was the resistance-network analogue method (Liebmann 1950), which had been employed previously to study the field distribution of symmetrical magnetic lenses (Liebmann and Grad 1951).

Typical results are given in figures 2–5. The radius of the larger of the two lens bores (figure 1) was taken as unit of measurement of length. In figure 2, the axial field strength H_z, expressed as a fraction of the field strength H_p in the parallel part of the pole piece gap, at great distance from the axis, is plotted for a series of lenses of equal gap width $S=R_2$ but varying radius R_1 of the smaller lens bore. The relative position of the pole faces remained constant during this series. It is seen that z_m, the position of the maximum H_m of the H_z curve, moves towards the pole face with the smaller bore as the bore radius R_1 is decreased to coincide with the left pole face for $R_1 \to 0$; a graph of $-z_m/R_2$ as a function of the bore ratio R_1/R_2 is given in figure 4. Of greater importance for the imaging properties of the lens is the fact that for decreasing value of R_1 the maximum value H_m increases, leading to a greater refracting power of the asymmetrical lens.

It is noteworthy that the value of H_m for all curves in figure 2 agrees very closely (to better than 1%) with the value of H_m as a function of S/D (see Liebmann and Grad 1951, fig. 7) if S/D is replaced by $S/(R_1+R_2)$ in this earlier graph.

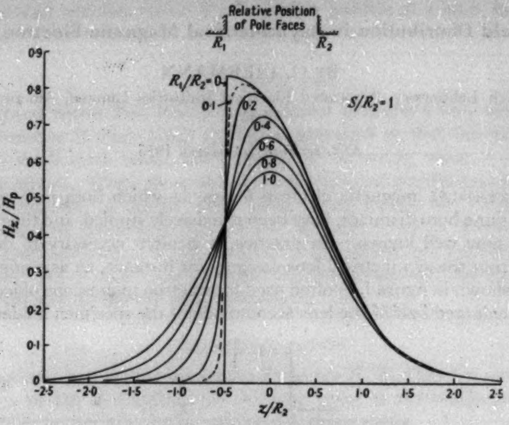


Figure 2.

The relative 'half-width' values a_1/R_2 and a_2/R_2 for the field distributions to the left and to the right of the field strength maxima H_m in figure 2, i.e. the distances from the maximum at which the field strength has fallen to $\frac{1}{2}H_m$, are plotted in figure 3 as functions of the ratio R_1/R_2 for this series of lenses. The ratio a_1/a_2 of the two half-widths may be considered as a useful measure of the electrical asymmetry, hence of the electron-optical asymmetry, of the lens whose mechanical, or design, asymmetry is characterized by the bore ratio R_1/R_2. Figure 4, in which a_1/a_2 is plotted as a function of the ratio R_1/R_2, shows that the deviation of the electrical asymmetry from the mechanical asymmetry is not very pronounced, the electrical asymmetry being always smaller than the mechanical asymmetry. The a_1/a_2 data of Dosse's and Lenz's work, mentioned before, come to lie exactly on the a_1/a_2 curve of figure 4, near its top end.

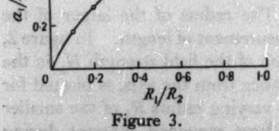


Figure 3.

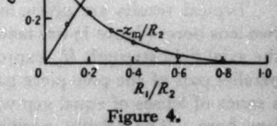


Figure 4.

Further measurements have shown that the electrical asymmetry parameter a_1/a_2 is a function of the bore ratio $R/_1R_2$ only and does not depend on the relative pole piece separation S/R_2. This is illustrated by the graphs of figure 5, which show the field distributions in three lenses of different gap widths S/R_2, from 0·5 to 2·0, the mechanical asymmetry being constant at $R_1/R_2=0{\cdot}3$. A ratio $a_1/a_2=0{\cdot}42\pm0{\cdot}01$ is found for all three lenses of figure 5. The data of this figure also confirm once more that the value of $H_\mathrm{m}/H_\mathrm{p}$ agrees with the value of H_0/H_p as function of S/D as given by Liebmann and Grad, if S/D is replaced by $S/(R_1+R_2)$.

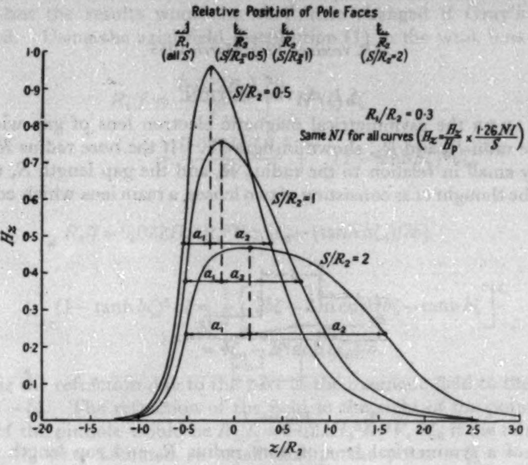


Figure 5.

ACKNOWLEDGMENT

The author is indebted to Dr. T. E. Allibone, the Director of the Laboratory, for his permission to publish this note.

REFERENCES

DOSSE, J., 1941, *Z. Phys.*, **117**, 437.
LENZ, F., 1950, *Optik*, **7**, 243.
LIEBMANN, G., 1950, *Brit. J. Appl. Phys.*, **1**, 92.
LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64**, 956.

The Magnetic Pinhole Electron Lens

By G. LIEBMANN
Research Laboratory, Associated Electrical Industries Limited, Aldermaston, Berks.

MS. received 18*th March* 1955

§ 1. Introduction

Consider the asymmetrical magnetic electron lens of gap width S and bore radii R_0 and R_1, shown in figure 1. If the bore radius R_1 becomes very small in relation to the radius R_0 and the gap length S, the lens of figure 1 can be thought of as consisting of two lenses, a main lens which corresponds to one half of a symmetrical lens of bore radius R_0 and gap length $2S$, and a 'pinhole' lens located in the plane of the pole face with the small bore, of radius R_1. The use of such a lens as an electron microscope objective was discussed by the author in a previous paper (Liebmann 1951), as it has the advantage of combining low spherical aberration with the requirement for a relatively small number only of ampere turns, provided the pinhole lens is located on the *incident* side of the electron beam. If, however, the pinhole lens is located at the emergent side, it may contribute substantially to the spherical aberration of the combined system. Such a situation is met for example, when one wishes to design a 'strong' lens whilst keeping the focus F (see figure 1) outside the magnetic field of the lens. It is therefore of some interest to have some quantitative information on the properties of the magnetic pinhole lens.

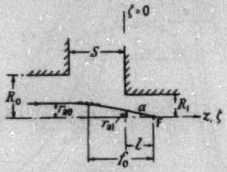


Figure 1.

§ 2. The Focal Length of the Pinhole Lens

Owing to the short distance z_0 within the pinhole lens over which the magnetic field falls to a very small value, it is permissible to apply the 'weak' lens formulae to it. Also, the axial field distribution within the pinhole can be described, according to Gray (1939) and Lenz (1950) with good approximation by the formula

$$H(\zeta)=\tfrac{1}{2}H_0(1-\tanh b\zeta). \qquad \ldots\ldots(1)$$

In equation (1) H_0 is the value of the axial field strength which would exist in the 'main' lens near the pole face containing the pinhole if the pinhole were absent (H_0 is equal to the maximum field strength along the axis of the symmetrical lens of radius R_0 and gap length $2S$), $\zeta=z/R_1$ the relative distance from the pole face containing the pinhole, and b a constant. According to Gray (1939) $b\simeq 1{\cdot}32$, whereas according to Lenz (1950) $b \simeq 1{\cdot}20$. Comparison with field measurements on asymmetrical lenses, as described in a preceding note (Liebmann 1955 a), showed that if S is of the same order of magnitude as R_0 the value of b given by Lenz gives a close match, whereas for small relative values of S the value of b given by Gray appears more favourable. In the subsequent discussion Lenz's value will be used, but the results would be very little changed if Gray's value were substituted. Using the axial field distribution (1) in the weak lens formula for R_1/f,

$$R_1/f = \frac{0{\cdot}022R_1^2}{V_r}\int_{-\zeta_0}^{\zeta_0} H^2(\zeta)\,d\zeta$$

$$= \frac{0{\cdot}022H_0^2R_1^2}{4V_r}\int_{-\zeta_0}^{\zeta_0}(1-\tanh b\zeta)^2\,d\zeta,$$

one finds

$$R_1/f = 0{\cdot}022H_0^2R_1^2V_r^{-1}[\zeta_0-(\tanh b\zeta_0)/2b],$$

since

$$\int_{-\zeta_0}^{\zeta_0}(1-\tanh b\zeta)^2\,d\zeta = \frac{1}{b}\Big[2b\zeta - 2\ln\cosh b\zeta - \tanh b\zeta\Big]_{-\zeta_0}^{\zeta_0}$$

$$= 4\zeta_0 - 2(\tanh b\zeta_0)/b.$$

This is the refraction due to the part of the magnetic field to the right of the point $\zeta = -\zeta_0$. The refraction of the field to the right of the point $-\zeta_0$ in the absence of the pinhole would be $R_1/f_1 = 0{\cdot}022H_0^2R_1^2V_r^{-1}\zeta_0$ if the relative pinhole radius R_1/R_0 were assumed so small that at the point $-\zeta_0 = -z_0/R_1$ (where $\tanh b\zeta_0 \simeq 1$, i.e. $\zeta_0 \simeq 2$, say) the field $H(-\zeta_0) \simeq H_0$. The presence of the pinhole lens would therefore lead to a loss of refraction

$$\Delta(R_1/f) = -0{\cdot}022H_0^2R_1^2V_r^{-1}(\tanh b\zeta_0)/2b.$$

Writing $k^2 = 0{\cdot}022H_0^2R_0^2V_r^{-1}$, and $\tanh b\zeta_0 \simeq 1$, this can be put in the simple form

$$\Delta(R_0/f) \simeq -0{\cdot}4k^2(R_1/R_0). \qquad \ldots\ldots(2)$$

This can be further simplified for lenses of the type considered in which $S \leqslant 2R_0$, and for weak and moderately strong excitation, as in this case $0{\cdot}4k^2 \simeq (R_0/f_0)R_0/(S+R_0)$ (see Liebmann and Grad 1951, fig. 15), f_0 being the focal length of the 'main' lens (remembering that this is one half of a symmetrical lens). Hence, the relative increase in focal length due to the presence of the pinhole is approximately

$$\Delta f/f_0 \simeq R_1/(S+R_0). \qquad \ldots\ldots(3)$$

The effect of relatively greater values S/R_0, or of strong lens excitation, would be to introduce a multiplying factor slightly greater than unity on the right-hand side of equation (3). This equation shows that the influence of the pinhole on the focal length is relatively small.

§ 3. SPHERICAL ABERRATION OF THE PINHOLE LENS

The spherical aberration constant C_{s0} of the 'main' lens was determined from the data on which were based earlier results published on symmetrical lenses (Liebmann and Grad 1951), by following the electron trajectories, computed for the earlier work just mentioned, up to the plane of symmetry of the symmetrical lens, and extrapolating the trajectory linearly from there. The spherical aberration constant C_{s0} can be represented in the same manner as in a recent paper on a 'unified' representation (Liebmann 1955 b), $C_{s0}/(S+R_0+R_1)$ being plotted as a function of $V_r/(nI)^2$. The resulting curve (figure 2) has the same character as the corresponding curve given in the earlier paper, but the C_{s0} values for the 'main' lens terminated by the pinhole lens are consistently lower, by a factor of about 0·75. The reason for this is that, unlike f_0, the constants C_s and C_{s0} are not in first approximation linear functions of $V_r/(nI)^2$. This favourable lower value of C_{s0} is, however, often vitiated by the influence of the spherical aberration of the pinhole lens, which must be added to C_{s0} to obtain the spherical aberration of the complete lens system, as shown in figure 1.

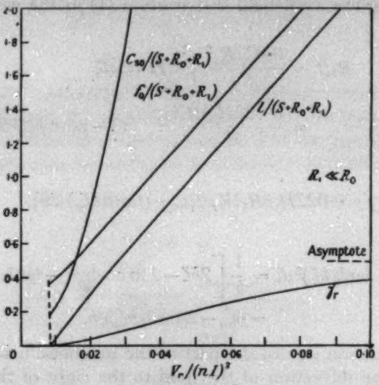


Figure 2.

The spherical aberration of the main lens is expressed most conveniently as the aberration angle $\delta\alpha_0$

$$\delta\alpha_0 = C_{s0}\alpha^3/f_0. \qquad \ldots\ldots(4)$$

The spherical aberration of the pinhole lens is then similarly described by the aberration angle $\delta\alpha_1 = \delta r_1/l$, where δr_1 is the radius of the disc of confusion in the gaussian image plane due to the aberration in the pinhole lens and l is the distance from the pinhole lens to the focus F (see figure 1), the lens aperture in the plane of the pinhole lens being $r_{a1} = \alpha l \simeq r_{a0} l/f_0$. The aberration angle $\delta\alpha_1$ can be obtained with sufficiently good approximation, for $R_1 \ll R_0$, from Scherzer's 'thin lens' formula

$$\delta\alpha_1 = 0{\cdot}011 r_{a1}{}^3 V_r{}^{-1} \int_{-\infty}^{\infty} H'^2(z)\,dz. \qquad \ldots\ldots(5)$$

Using the field distribution given by equation (1), the integral in equation (5) becomes:

$$\int_{-\infty}^{\infty} H'^2(z)\,dz = (b^2 H_0{}^2/4R_1)\int_{-\infty}^{\infty} \operatorname{sech}^4(b\zeta)\,d\zeta$$

$$= (bH_0{}^2/4R_1)\Big[\tfrac{1}{3}\tanh(b\zeta)(2+\operatorname{sech}^2(b\zeta))\Big]_{-\infty}^{\infty} \simeq 0{\cdot}4 H_0{}^2/R_1.$$

Hence $\delta\alpha_1 = 0{\cdot}044 H_0{}^2 V_r{}^{-1} R_1{}^{-1} \alpha^3 l^3$.

The relative contribution of the pinhole lens to the spherical aberration of the whole system can be conveniently expressed as the ratio $\delta\alpha_1/\delta\alpha_0$

$$\delta\alpha_1/\delta\alpha_0 = 0{\cdot}044 H_0^2 l^3 f_0/(V_r C_{s0} R_1). \qquad \ldots\ldots(6)$$

This can be rewritten as

$$\frac{\delta\alpha_1}{\delta\alpha_0} = 0{\cdot}2 k_0^2 \left(\frac{l}{R_0}\right)^3 \frac{f_0}{C_{s0}} \frac{R_0}{R_1} = \gamma \frac{R_0}{R_1}. \qquad \ldots\ldots(7)$$

Whilst the factor γ in (7) is a function of k_0^2, one knows that $\gamma \to 0$ when $l \to 0$, and that $\gamma = \text{constant}$ if the main lens is very weak, as then $l/R_0 \propto k_0^{-2}$ and $f_0/C_{s0} \propto k_0^4$. This constant value of γ for the weak lens condition is $\gamma \simeq 0{\cdot}5$.

The constant γ as a function of $V_r/(nI)^2$ is shown in figure 2, together with $C_{s0}/(S+R_0+R_1)$, $f_0/(S+R_0+R_1)$ and $l/(S+R_0+R_1)$. In this representation, only minor deviations occur from these curves for lenses where $S/R_0 \ll 1$, or $S/R_0 \gg 1$, as in the case of the 'unified' focal length curves (Liebmann 1955 b).

Relation (7), in connection with figure 2, shows that mostly the pinhole lens makes a considerable contribution to the total spherical aberration, as $R_0/R_1 \gg 1$; for relatively weak lenses, the pinhole increases the spherical aberration by a large factor.

ACKNOWLEDGMENTS

The author wishes to thank Mr. M. E. Haine for helpful discussions, and Dr. T. E. Allibone, Director of the Laboratory, for permission to publish this paper.

REFERENCES

GRAY, F., 1939, *Bell Syst. Tech. J.*, **18**, 1.
LENZ, F., 1950, *Ann. Phys., Lpz.*, (6), **8**, 124.
LIEBMANN, G., 1951, *Proc. Phys. Soc.* B., **64**, 972; 1955 a, *Ibid.*, **68**, 679; 1955 b, *Ibid.*, **68**, in the press.
LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64**, 956.

Radio-Carbon Content and Delayed Boiling of Liquids

By C. DODD
University College, London

Communicated by H. S. W. Massey; MS. received 11*th January* 1955 *and in final form* 5*th April* 1955

§ 1. Introduction

A LIQUID subjected to an external pressure equal to or less than the saturation vapour pressure of the liquid at that temperature boils, but the boiling may be immediate or delayed. For boiling to begin, conditions must be such that a bubble of microscopic size formed within the liquid can increase in size continuously. Its internal pressure must therefore exceed the sum of the hydrostatic pressure and the pressure due to liquid surface tension both of which tend to collapse the bubble. Such bubbles are being continually formed in the liquid by random molecular movement but they are usually so small that the pressure due to surface tension, which is inversely proportional to the radius of the bubble, is sufficient to cause their immediate collapse. Boiling is of course immediate should there be present solid nuclei upon which bubbles of finite size can form, but in the case of a clean liquid contained in a smooth vessel the boiling will be delayed. In the absence of any electric charge on the bubble, boiling will then not occur until thermal fluctuations produce an internal pressure in the minute bubble great enough to enable it to grow to visible size. At fairly low degrees of superheating the probability of this occurring is so low that an appreciable delay occurs before the liquid begins to boil.

Should however the bubble acquire an electric charge on its formation, an extra pressure tending to promote growth of the bubble exists and this pressure increases inversely as the fourth power of the radius of the bubble. A bubble of small size which would otherwise be collapsed by the surface tension effect may thus be able to grow because of the charge it carries. There are two possible ways in which a bubble can acquire such a charge. The obvious way is by ionization through an external agent such as a cosmic ray or a radioactive source. The other way is by ionization occurring internally in the liquid. Such ionization would be expected in all liquids containing radio-carbon atoms since ^{14}C disintegrates with a half-life of about 5000 years giving a β-particle of energy 155 kv. The object of these experiments is to investigate the effect of these ionizing sources on the time delay occurring before a superheated liquid boils.

§ 2. Apparatus

The liquid under test is contained in a thick-walled glass bulb of volume about 3 cm^3 and is immersed in a bath of glycerine whose temperature could be maintained at any value up to 150°c (figure 1). Pressures up to 35 atmospheres are applied to the working liquid by connecting the exit tube of the pressure chamber to a cylinder of compressed air, the pressure being transmitted to the liquid via a rubber diaphragm faced with a thin sheet of P.T.F.E. to avoid contamination of the liquid. Leakage of liquid past the sides of the capillary is prevented by O rings compressed by brass washers. The experiment consisted of maintaining the temperature of the liquid at a constant value well above the normal boiling point and applying sufficient pressure to prevent the liquid from boiling. By opening a valve this pressure could be rapidly reduced to atmospheric. In general the liquid then remained undisturbed for some time before violent boiling occurred and this delay time was measured by a stopwatch. The bulb is refilled ready for the next expansion by application of pressure to the diaphragm, this resetting taking only a few seconds.

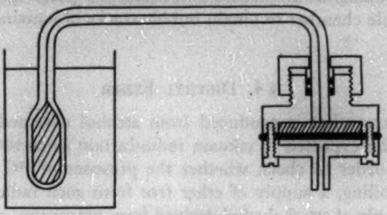


Figure 1.

Since at any temperature the delay time before the onset of boiling is very variable, it was necessary to find the mean value of the delay time for 50 expansions at each temperature.

§ 3. PENTANE

Pentane is derived from petroleum sources of so great an age that no radiocarbon remains. The variation of delay time with temperature is shown by the full curve in figure 2. Although the normal boiling point is 36°c, even at 120°c an average delay time of over half a minute elapses before the onset of boiling

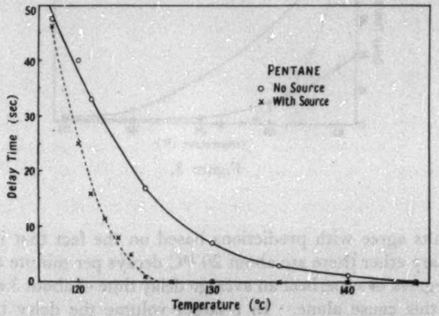


Figure 2.

since the only ionizing agents are cosmic rays and the probability of these producing a large charge on a bubble is small. At higher temperatures the surface tension is reduced and not only is the probability of thermal fluctuations producing a bubble large enough to grow increased, but also a much smaller ionization charge is required for bubble growth. Both effects give rise to a big decrease in delay time until at 145°c boiling always occurs almost immediately.

On placing a 100 microcurie ^{60}Co source near to the liquid chamber, the delay time follows the course shown by the broken curve in figure 2. In the region of 120°c there is a finite probability of a bubble acquiring a sufficiently large charge by ionization to enable it to grow, so the average delay time is much less with the source present, and at all temperatures above 125°c this probability is so great that boiling always occurs immediately. Glaser (1953) has used this effect in his bubble chamber to obtain bubble tracks of ionizing particles.

§ 4. DIETHYL ETHER

In this country ether is produced from alcohol obtained by fermentation and thus would be expected to contain radio-carbon since the half-life is some 5000 years. In order to check whether the presence of ^{14}C atoms affects the delay time for boiling, a supply of ether free from such radioactive atoms was required. A supply of ethyl alcohol derived from petroleum, and thus free from radio-carbon, was obtained and a sample of diethyl ether prepared from it. The same test bulb was filled first with ordinary ether and then with the specially prepared ether and the delay times for boiling obtained for a series of temperatures. The curves are quite different, the delays for ordinary ether being always much less than for the special ether free from radio-carbon, and conform closely to the delays obtained using special ether in the presence of the ^{60}Co source (figure 3).

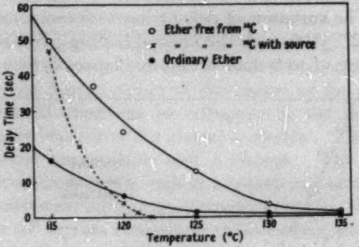


Figure 3.

These results agree with predictions based on the fact that in a volume of 3 cm^3 of ordinary ether there are about 20 ^{14}C decays per minute so that at sufficiently high degrees of superheat an average delay time of about 3 seconds should occur due to this cause alone. In a larger volume the delay time would be proportionately less and for the large volume bubble chambers proposed by Glaser and Rahm (1955), several decays can be expected in the few milliseconds during which such chambers remain sensitive. For this reason it is desirable to avoid the use of working liquids containing ^{14}C if large bubble chambers are to be operated successfully and with low background.

Acknowledgments

The author would like to express his thanks to Dr. Donald Glaser for his advice on the technique for successful operation of his bubble chamber and to Professor C. K. Ingold for granting the facilities in his Department of Chemistry for the preparation of the ether free from radio-carbon.

References

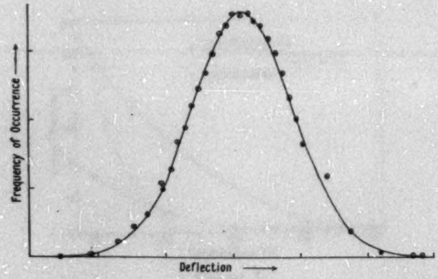
Glaser, D. A., 1953, *Phys. Rev.*, **91**, 763.
Glaser, D. A., and Rahm, D. C., 1955, *Phys. Rev.*, **97**, 474.

LETTERS TO THE EDITOR

Distribution Function of Semiconductor Noise

It is commonly assumed that the gaussian distribution of instantaneous values which applies to 'white noise' will also apply to current noise in semiconductors, and recent theoretical investigations by Burgess (1954) and by the writer (1954) have shown that the recombination conditions which are likely to apply in a semiconductor lead to a very nearly gaussian distribution for the instantaneous numbers of conduction electrons (or holes) provided the numbers are large. But in view of the difficulty of explaining the inverse-frequency power spectrum in terms of the usual models of the conduction mechanism, it seemed worth while to confirm experimentally that there was no substantial departure from gaussian distribution of noise voltage in a typical specimen which showed the inverse-frequency spectrum of current noise.

The sample examined was a carbon film of surface 1·9 cm long, 1·5 cm wide and having a resistance of 16 000 ohms, which had been prepared for the writer by the Welwyn Electrical Laboratories Ltd. Its power spectrum in the range covered by the experimental equipment, namely 40 to 6000 c/s,† has been examined by Mr. R. J. K. Ford and found to follow the law f^{-n} where $n=1\pm0{\cdot}02$. The amplitude distribution was explored by displaying the noise voltage on a cathode-ray tube with a time base of arbitrary speed and measuring the mean intensity at various amplitude levels by means of a slit and photomultiplier. The time base was used (rather than allow the signal to appear as a single vertical line of varying intensity) to reduce the risk of saturation of the fluorescent screen; and the satisfactory operation of the complete equipment was verified by comparing measured and theoretical distributions with the input first consisting of a sine wave and then consisting of shot noise from a temperature-limited diode. The results for the carbon film are shown in the figure, where the experimental points are indicated by



---

† This bandwidth is many times that of the indicating device, so that requirements of the type discussed by Arthur (1952) are satisfied.

circles and the line shows a gaussian curve fitted to these points. There is no departure from the gaussian distribution which is significant within the accuracy of this simple method of measurement, which does not, however, justify the precise significance tests proposed by Siegert (1952).

The writer wishes to record his thanks to Mr. R. J. K. Ford, who built the equipment for generating and amplifying the current noise, and to Professor A. Tustin who provided facilities for this work in the Electrical Engineering Department.

Electrical Engineering Department,
University of Birmingham.
27th May 1955.

D. A. BELL.

ARTHUR, G. R., 1952, *J. Appl. Phys.*, **23**, 1143.
BELL, D. A., 1954, *Proc. Phys. Soc.* B, **67**, 227.
BURGESS, R. E., 1954, *Physica*, **20**, 1007.
SIEGERT, A. J. F., 1952, *J. Appl. Phys.*, **23**, 737.

REVIEWS OF BOOKS

The Properties of Glass, by G. W. Morey. 2nd Edn. Pp. 591. (New York: Reinhold Publishing Corporation, 1954).

When the first edition of Morey's well-known treatise appeared in 1938, this reviewer felt that it must stand for all time as a model for authors wishing to combine a maximum of empirical information with a minimum of critical comment and discussion. In the face of some recent competition a second edition now appears which combines a still greater quantity of empirical information with almost the same amount of—and almost the same—comment. The appeal is still to the glass technologist who has to get on with the job, and the discussion of principles is cursory and faint-hearted, with liberal use of quotation marks to indicate that views expressed are not necessarily those of the author. In spite of claims made in the preface, recent work on fundamental properties of glass has been mainly ignored except for the insertion of additional references in footnotes. This book of nearly 600 pages on the properties of glass does not answer the first questions which would be asked by an intelligent physicist newly introduced to the field. An unintelligent physicist would not guess from reading this book that there were any questions to be asked.

G. O. J.

Physicochemical Calculations, by E. A. Guggenheim and J. E. Prue. Pp. xii + 492. (Amsterdam: North Holland Publishing Company, 1955.) 53*s*.

This masterly collection of 171 worked problems based on published experimental data will be a boon to teachers and students of physical chemistry. The principle of the book is that textbooks are not enough, understanding only comes by applying, or, as here, studying the application of textbook knowledge to individual problems. A wide range of subjects in chemical physics and physical chemistry is covered with characteristic precision and clarity. Many of the problems are of direct interest to physicists; those on more specifically chemical subjects—chemical kinetics and thermodynamics, electrolyte solutions, surface chemistry, etc.—will be useful to physicists who require more than a textbook acquaintance with these subjects. The choice necessarily reflects the interests of the authors. A noteworthy omission is that despite the excellent and full collection illustrating the thermodynamic uses of reversible electrodes, none are included on electrode kinetics to provoke the student to consider how reversible potentials are established.

The discussion sections are illuminating but occasionally Olympian in shortness. Some might have been expanded into the blank spaces at the end of the problems (each problem starts elegantly—and very rightly—on a new page) with advantage to many readers and without loss of edge. For example, in problem 83 most readers would have welcomed a discussion of the advantages and snares of estimating the error in the slope of lines by numerical methods instead of by the method used here, judgment by eye. In problem 122 the impression will be left on the guileless that in estimating dy/dx ($=$constant) given y for three equally spaced values of x the middle value contributes nothing (it gives, of course, a rough measure of error). In problem 14 the effect on the estimate of Avogadro's number of taking the modern value of the disintegration rate of radium instead of Rutherford and Geiger's value should have been mentioned.

L. YOUNG.

The Illumination and Polarisation of the Sunlit Sky on Rayleigh Scattering, by S. CHANDRASEKHAR and DONNA D. ELBERT. Pp. 86. (Philadelphia: *Transactions of the American Philosophical Society*, Vol. 44, Pt. 6, 643, 1954.) $ 2.00.

The problem of the reflection and transmission of a parallel beam of light in a given state of polarization by a thick layer of atmosphere showing Rayleigh scattering has been solved by the senior author (*Radiative Transfer*, S. Chandrasekhar, Oxford: University Press, 1950). The present memoir gives an outline of the solution together with tables of the basic functions involved. The introduction also contains sets of curves showing the distribution of polarization for various cases. H. H. HOPKINS.

Reihenentwicklungen in der mathematischen Physik, by J. LENSE. Pp. 216. (Berlin: de Gruyter, 1953.) DM. 26.

The use of expansions in terms of orthogonal functions is now such a commonplace in mathematical physics that an account of both these and related expansions and functions cannot fail to be of interest to a large number of physicists. Professor Lense's book, the first edition of which appeared more than twenty years ago, deals with the properties of Fourier series, and the polynomials associated with the names of Bernoulli, Laguerre, Hermite and Tschebyschef, together with extensive accounts of Bessel functions, spherical harmonics and Lamé functions. Complex theory is only used in connection with Bessel functions; for the rest a real variable is assumed. In addition to the treatment of specific functions, a number of techniques is also described, including, for example, the method of steepest descent and a discussion of 'completeness' as applied to orthogonal functions. Considered as a general view of this branch of mathematics, it is not easy to imagine more material being described so well in so short a space. H. H. HOPKINS.

On the Origin of the Solar System, by H. ALFVÉN. Pp. 194. (Oxford: University Press, 1954.) 30*s*.

In this volume the hypothesis that electromagnetic forces constitute one of the key processes in the formation of the solar system is investigated. The behaviour of an ionized mixture of gases under the combined influence of a magnetic field and the gravitation of the sun is first discussed. The differences of ionization potential of the constituents afford some scope for the separation out of different elements at different distances from the sun to give regions of differing compositions. So long as the full implications of the assumed magnetic fields are not gone into too deeply the process holds out the possibility of the gases being stopped in their infall at certain distances. Angular momentum can then possibly be transferred to the material from the assumed rapidly rotating sun by further electromagnetic action, and the rings of gas so arrested and accelerated into orbital motion then condense into droplets, which go on to grow into planets in the general manner recently advocated by a number of writers on this problem.

The author regards it as an advantage that no unusual or rare event has to be postulated but only the interaction of the sun with interstellar material already known to exist, but such philosophical discussion as is embarked upon in the earlier pages savours too much of the nonsense that so many astronomers are, alas, imbued with and regard as incontrovertible commonsense to be worth inclusion in the volume or contradiction here. The main misgiving that this reviewer felt throughout the main body of the book was whether in fact any *form* of magnetic field that could reasonably be postulated for the sun (a dipole strength 10^5 times the present value being regarded as allowable) would operate in the way implied. There must always be questions in the background of how the lines of magnetic force are originated, and although one is here considering them in (effectively) empty space they must have their roots or origin somewhere, so that it is important to be sure that any postulated field is not implying huge blocks of immovable permanent magnets approaching galactic dimensions in size tacitly hidden away, or comparable current systems, for instance. It therefore came as a surprise to find explicit reference to this kind of difficulty only in the closing pages of the book (p. 180) where it is suddenly remarked that *the lines of force must not go directly from the cloud to the central body*, whereas throughout most of the preceding 179 pages the reader has been allowed to think that the sun's dipole field, which has this inconvenient property, is suitable (at however enhanced a strength). Full attention to the difficulty might of course have involved a complete reconsideration of all that had gone before, and it is a moral problem for an author making such a discovery late in a piece of work to decide what he shall do about it.

There is little doubt that whatever processes in astrophysics turn out to be principally electromagnetic, Professor Alfvén will have been among the first to have suggested them, but this reviewer is doubtful of the extent to which the present volume provides convincing evidence that they include the formation of the solar system, at any rate by the mechanism considered. Professor Alfvén is always interesting and stimulating, and the student of magneto-hydrodynamics can learn much from his works, so that the value of his book is perhaps to be measured more from this point of view than in terms of any definite finished achievement in the field of cosmogony. Professor Alfvén is a pioneer in a realm that abounds with problems of satanic difficulty with which someone has got to be prepared to make a start, and we must all congratulate him on his courageous explorations. R. A. LYTTLETON.

Elektronische Halbleiter: eine Einführung in die Physik der Gleichrichter und Transistoren, by E. SPENKE. Pp. xx + 379. (Berlin, Göttengen, Heidelberg: Springer, 1955.) Price DM. 34.50.

This book is a welcome addition to the still rather limited number of general texts on the fundamentals of solid state physics. While primarily intended for those concerned with crystal rectifiers and transistors, it seeks to provide the reader with a background of knowledge which he can apply to any type of solid state problem.

The first five chapters give a survey of the main phenomena and basic concepts, leading to a discussion of the various types of rectifier and transistor. The treatment here of band structure, electrons and holes and effective mass will provide an excellent introduction for those new to the subject. The sections on contacts and on the devices themselves are clear and thorough also. Ionic crystals are included in these chapters, and the section on defects, factors affecting their concentration, and the application of the laws of mass action is particularly welcome. These matters are rather inadequately treated in most texts available so far in English.

The final five chapters, comprising more than half the book, deal with the foundations of semiconductor physics. There is first a formal treatment of the band theory, comparing the Bloch and Brillouin approximations, and giving physical interpretations. This leads to detailed discussions of the motion of the electron in an applied field, the effective mass, the Zener effect, scattering processes, and so to the theory of conductivity. This is perhaps the most important part of the book and will repay careful study by students of the solid state at all levels.

The Fermi statistics are then discussed. This is a less thorough treatment, and in particular there is no mention of the effects arising in the statistics when the spin of the electron trapped at donor levels is considered. This omission is disappointing, especially since the matter is referred to in Shockley's book only in the problems at the ends of chapters 10 and 16, and there is evidence that there is some lack of general understanding.

The next section deals with the equilibrium conditions in an impurity semiconductor, the application of the law of mass action, and the lifetimes of the neutral and ionized donors. The final chapter discusses surface phenomena, including those at the contact between a metal and a semiconductor. This chapter gives a thorough treatment of the basic features, and will be valuable since these are still commonly misunderstood, not only by beginners.

The scope of such a volume as this has of course to be limited; the author lists in his introduction numerous subjects he has not been able to include; no doubt others will occur to readers. The present writer would have been pleased to see a fuller discussion of the Hall and related effects, and a development of the section on surfaces to include inversion layers, surface conductivity, surface recombination, and some account of the effects of adsorbed layers. These features, however, would considerably enlarge a book already generous in subject matter.

The production is good, there are very few minor errors, the references are adequate, and in all respects the book can be thoroughly recommended.

D. A. WRIGHT.

CONTENTS OF SECTION A

PAGE

Dr. A. H. de Borde and Prof. H. S. W. Massey. The Collisions of Nucleons with Deuterons 769

Dr. G. M. Griffiths and Prof. J. B. Warren. The Gamma Radiation from the Reaction $D(p, \gamma)^3He$ 781

Mr. D. A. Eyeions, Dr. B. G. Owen, Mr. B. T. Price and Prof. J. G. Wilson. The Ionization Loss of Relativistic μ-Mesons in Neon 793

Mr. D. West, Mr. R. Batchelor and Mr. E. F. Bradley. The Detection of Mesic X-rays with Pressurized Proportional Counters using Cosmic Ray μ-Mesons stopped in Carbon 801

Mr. R. W. McCarroll and Mr. M. R. C. McDowell. Properties of the Hydrogen Molecular Ion: VI—The Total Electronic Orbital Angular Momentum of the $1s\sigma_g$, $2s\sigma_g$ and $3s\sigma_g$ States 810

Mr. E. Holøien and Mr. J. Midtdal. On a Metastable Energy State of the Negative Helium Ion 815

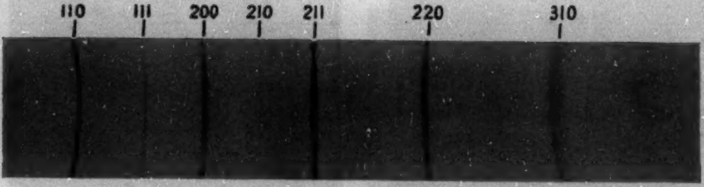
Mr. P. G. Dodsworth and Dr. R. F. Barrow. The Triplet Band Systems of Aluminium Monofluoride 824

Dr. E. P. George and Dr. J. Evans. Further Observations of Cosmic-Ray Events in Nuclear Emulsions Exposed Below Ground 829

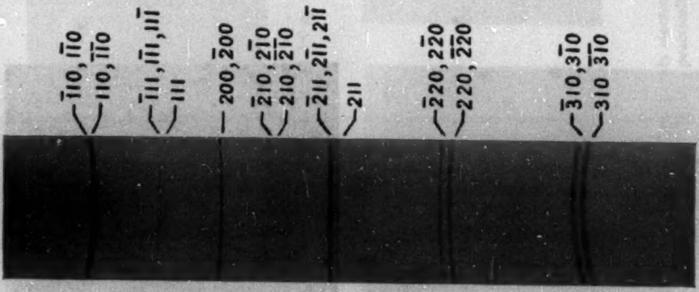
Research Notes:

Mr. B. Ledley and Dr. L. Riddiford. Determination of Proton Synchrotron Beam Size with a Crystal Scintillator 836

Contents of Section B 840

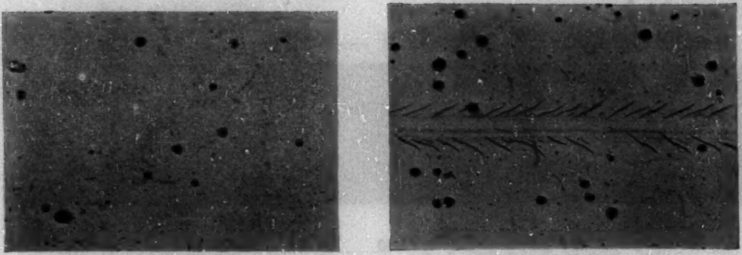

110
111
200
210
211
220
310

(*a*) La Mn $O_{3\cdot10}$ Cubic

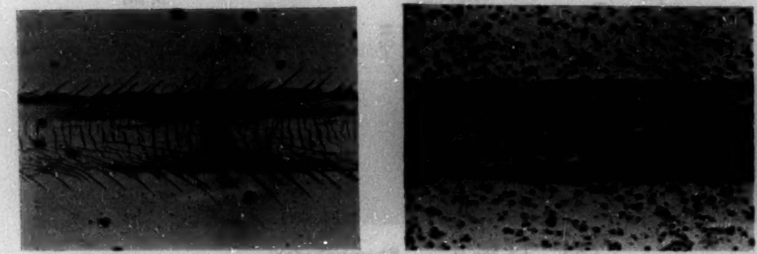

$\bar{1}10,1\bar{1}0$
$110,\bar{1}\bar{1}0$
$\bar{1}11,1\bar{1}1,11\bar{1}$
111
$200,\bar{2}00$
$\bar{2}10,2\bar{1}0$
$210,\bar{2}\bar{1}0$
$\bar{2}11,2\bar{1}1,21\bar{1}$
211
$\bar{2}20,2\bar{2}0$
$220,\bar{2}\bar{2}0$
$\bar{3}10,3\bar{1}0$
$310\ \bar{3}\bar{1}0$

(*b*) La Mn $O_{3\cdot15}$ Rhombohedral

Figure 1. X-ray diffraction patterns of samples having different structures.

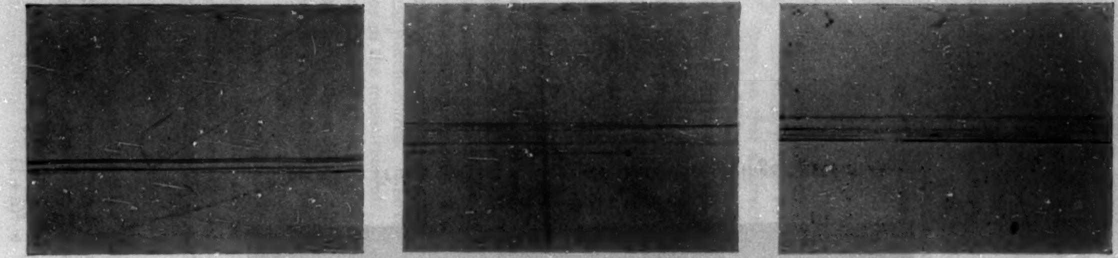


A. Load 2 g, ×667 B. Load 20 g, ×133



C. Load 200 g, ×67 D. Load 2000 g, ×27

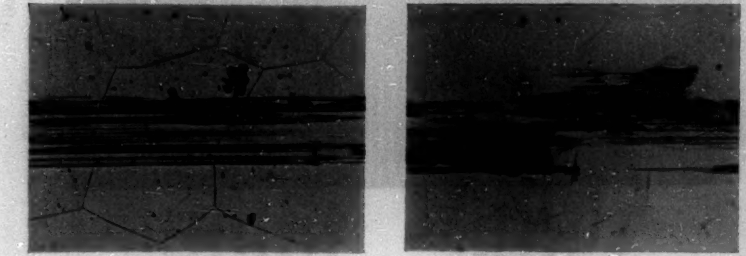
Figure 5. Stages in the breakdown of an anodized layer on aluminium. Direction of sliding from left to right.



A. Load 0·2 g, ×1067

B. Load 2·0 g, ×1067

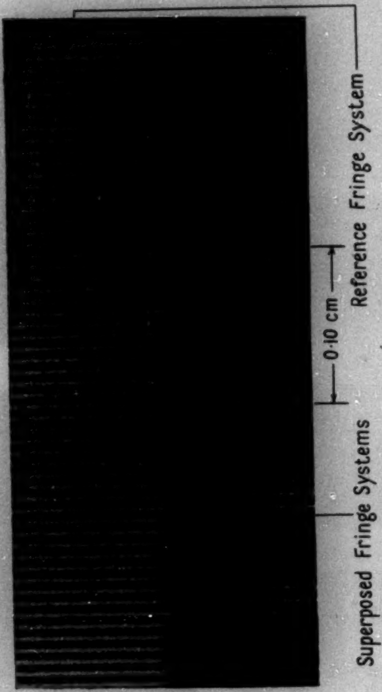
C. Load 20 g, ×220



D. Load 100 g, ×220

E. Load 200 g, ×100

Figure 6. Tracks on the surface of an annealed, abraded and lightly electropolished nickel block. Contact resistance measurements showed the oxide breakdown was practically complete at 2 g load.


Reference Fringe System
0·10 cm
Superposed Fringe Systems

Measurements on p–n Junctions in Lead Sulphide

By T. S. MOSS

Radio Department, Royal Aircraft Establishment, Farnborough, Hants.

MS. received 4th March 1955 *and in final form* 21st *April* 1955

Abstract. From measurements of the characteristics of several natural p–n junctions the intrinsic resistivity of lead sulphide is calculated to be 3·1 Ω cm. Hence at room temperature the intrinsic carrier concentration is $2{\cdot}7 \times 10^{15}$ electrons or holes per cm^3 and the width of the forbidden zone 0·40 ev.

---

§ 1. Introduction

So far it has not been possible to obtain single crystals of PbS which are sufficiently pure to be in the range of intrinsic conductivity at room temperature, so that a direct measurement of the intrinsic resistivity is not yet possible. For this reason indirect measurements of this parameter are of interest. Also such widely varied values of the thermal activation energy have been deduced from Hall effect measurements that an independent method of obtaining the gap width is important. As natural p–n junctions have now been detected in PbSe by Goldberg and Mitchell (1954), the method here described may also be applicable to this material.

§ 2. Experimental

Several p–n junctions were located in crystals of Joplin galena, and potentiometric measurements made of their characteristics. Small samples were used in order to increase the likelihood of uniformity of the junction, the cross-sectional areas ranging from 10^{-3} to $10^{-2}\,\mathrm{cm}^2$. As the junctions were never simple planes perpendicular to the length of the specimens it was necessary to trace the contour of the junction roughly on all four faces of the specimen in order to estimate the actual junction area. The estimation of this area is probably the greatest source of error in the measurements. Figure 1 shows the rectification ratio of one of the best junctions (junction b of the table). It will be observed that the plot is linear up to about 30 mv, the slope of the straight line drawn being $\boldsymbol{e}/\boldsymbol{k}T$, i.e. the junction is theoretically perfect for low applied fields, and gives the correct value for the electron charge, unlike point contact rectifiers.

It was not possible to plot the potential distribution over the sharper junctions, but for a rather diffuse one which gave a rectification ratio of 1·75 : 1 at $\boldsymbol{e}V=\boldsymbol{k}T$, the plot of figure 2 was obtained. It will be seen that 60% of the potential change occurs in a distance of 100 μ, and 85% in 200 μ.

§ 3. Theory

According to Shockley (1950) the current density through a perfect p–n junction (i.e. one in which the transition region is so narrow that negligible recombination takes place therein) is

$$i=\frac{b\sigma_i^2}{(1+b)^2}\frac{\boldsymbol{k}T}{\boldsymbol{e}}(\exp \boldsymbol{e}V/\boldsymbol{k}T-1)\left(\frac{1}{\sigma_n L_p}+\frac{1}{\sigma_p L_n}\right)$$

where $b=\mu_n/\mu_p$ = mobility ratio, σ_i = intrinsic conductivity, σ_n, σ_p = conductivity of the n and p regions respectively, L_n = diffusion length of the electrons in the p-region, L_p = diffusion length of the holes in the n-region.

Hence the rectification ratio for a voltage $\pm V$ across the junction is $i_f/i_r = \exp \boldsymbol{e}V/\boldsymbol{k}T$ where i_f and i_r are the currents in the forward and reverse directions respectively. A perfect junction will thus give a rectification ratio of e : 1 for an applied voltage of $\boldsymbol{k}T/\boldsymbol{e}$, i.e. 25 mv at 293°K. The observed rectification ratio at this voltage for an actual junction can be taken as a measure of the degree of perfection of the junction.

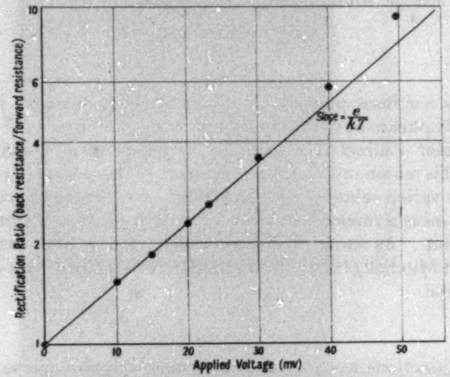


Figure 1. Rectification characteristic of PbS p–n junction.

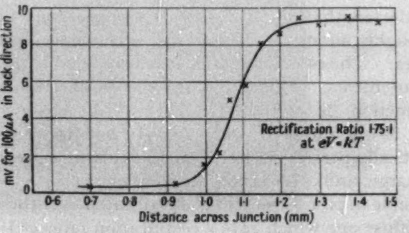


Figure 2. Potential distribution across diffuse p–n junction in PbS.

The most difficult factors to evaluate in the equations are normally the $L_p\sigma_n$ terms. However it has been shown (Moss 1953) that for different specimens of PbS the carrier lifetime τ is proportional to the square of the specimen resistivity. Hence $\sigma_1\tau_1^{1/2} = \sigma_2\tau_2^{1/2}$ = constant.

From the measurements quoted the average value of $\sigma^2\tau$ was found to be 2×10^{-5} sec Ω^{-2} cm^{-2}, there being no significant difference between p- and n-type specimens. Hence for the n-region $\sigma_n L_p D_p^{-1/2} = (\sigma\tau^{1/2})_n = (2\times10^{-5})^{1/2}$, and for the p-region $\sigma_p L_n D_n^{-1/2} = (\sigma\tau^{1/2})_p = (2\times10^{-5})^{1/2}$. Rewriting in terms of the combined mobility $\mu_p + \mu_n = (\boldsymbol{e}/\boldsymbol{k}T)(1+b^{-1})D_n$ we obtain

$$i = \sigma_1^2 \left(\frac{\boldsymbol{k}T}{\boldsymbol{e}}\right)^{1/2} (\exp \boldsymbol{e}V/\boldsymbol{k}T - 1) \frac{(b+b^{1/2})}{(1+b)^{3/2}} \frac{1}{\{(\mu_p+\mu_n)2\times10^{-5}\}^{1/2}}.$$

The current density in the forward direction at $\boldsymbol{e}V/\boldsymbol{k}T = 1$ is given by

$$I_f = \frac{62\sigma_1^2}{(\mu_p+\mu_n)^{1/2}} \frac{b+b^{1/2}}{(1+b)^{3/2}} \text{ amp cm}^{-2} \text{ at } 293^\circ\text{K}.$$

The mobility ratio is not well known for PbS. The ratios of actual mobilities measured on different p and n type samples give $b \sim 1$, although it is known from the sign of the Hall effect in the intrinsic range that $\mu_p > \mu_n$. However the required function of b is not critical, varying only from 0·70 to 0·65 for $b=1$ to $b=2$, so that the function may be taken as 0·68 with little possible error.

From Hall effect data which are available for natural crystals it appears that the best average value of $\mu_p + \mu_n$ is 800 cm^2 v^{-1} sec^{-1} (see Moss 1953, Scanlon and Brebrick 1954). Hence $I_f = 1{\cdot}5\,\sigma_1^2$ or $\rho_1^2 = 1{\cdot}5/I_f$.

Thus from measurements of the current density in the easy flow direction at an applied voltage of $\boldsymbol{k}T/\boldsymbol{e}$ (25 mv at 293°K) the intrinsic resistivity of the specimen can be determined.

The table summarizes the data obtained for the seven best junctions (i.e. those with highest rectification ratios).

Junction	a	b	c	d	e	f	g
Rectification at $\boldsymbol{e}V = \boldsymbol{k}T$	2·9 : 1	2·75 : 1	2·55 : 1	2·5 : 1	2·5 : 1	2·4 : 1	2·3 : 1
Calculated ρ_1 (Ω cm)	3·1	3·4	2·3	2·4	3·2	3·0	4·4

The average resistivity is thus $\rho_1 = 3{\cdot}1$ Ω cm.

The main sources of error in the estimation of τ (as a function of σ) and in the junction area. As $\rho_1 \propto \tau^{1/2}$ the uncertainty in ρ_1 due to inaccuracy in τ is probably of the order of ±20%. The dependence on junction area is $\rho_1 \propto (\text{area})^{1/2}$ so that the uncertainty due to this factor is also of the order of ±20%.

Allowing for these errors we would expect a scatter of results between 2·2 and 4·4 Ω cm, which is very nearly the range of values quoted in the table.

From the above value of ρ_1 we can calculate the intrinsic carrier concentration n_1 from the relation $1/\rho_1 = n_1 \boldsymbol{e}(\mu_p+\mu_n)$, which gives $n_1 = 2{\cdot}7\times10^{15}$ cm^{-3}. We can also deduce that the intrinsic lifetime $\tau_1 \sim 180\,\mu$sec, or perhaps twice this depending on the precise recombination mechanism.

In terms of the activation energy E, the intrinsic carrier concentration at room temperature is given by

$$n_1 = 2{\cdot}4\times10^{19} \frac{(m_e m_h)^{3/4}}{m^{3/2}} \exp(-E/2\boldsymbol{k}T),$$

where m_e and m_h are the effective masses of electrons and holes respectively. It has been shown theoretically by Bell *et al.* (1953) that $m_h \sim \frac{1}{2}m$. With $m_e m_h = \frac{1}{4}m^2$ we obtain $E = 0{\cdot}40$ ev.†

§ 4. Conclusions

It is concluded that for natural galena crystals the intrinsic resistivity at room temperature lies between 2·2 and 4·4 Ω cm, the most probable value being $\rho_1 = 3{\cdot}1$ Ω cm, which is very near to the figure obtained by extrapolating the data of

---

† It should be noted that this is the room temperature value of the gap width, not the value at absolute zero.

Scanlon and Brebrick (1954) to room temperature. From this latter value the forbidden gap width is calculated to be 0·40 ev, in close agreement with the now well established value of optical activation energy (see Moss 1953).

ACKNOWLEDGMENT

Acknowledgment is due to the Chief Scientist, Ministry of Supply, for permission to publish this article.

REFERENCES

BELL, D. G., HUM, D. M., PINCHERLE, L., SCIAMA, D. W., and WOODWARD, P. M., 1953, *Proc. Roy. Soc.* A, **217,** 71.
GOLDBERG, A. E., and MITCHELL, G. R., 1954, *J. Chem. Phys.*, **22,** 220.
MOSS, T. S., 1953, *Proc. Phys. Soc.* B, **66,** 993.
SCANLON, W., and BREBRICK, R. F., 1954, *Phys. Rev.*, **94,** 1430.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : van Nostrand).

Fluorescence Excitation Spectra and Quantum Efficiencies of Organic Crystals

By G. T. WRIGHT†

Physics Department, Rhodes University, Grahamstown, South Africa

MS. received 13*th April* 1955 *and in final form* 11*th May* 1955

Abstract. Fluorescence excitation spectra are measured for some organic crystals using excitation wavelengths down to 2200 Å. The fluorescence quantum efficiency of crystalline *p*-terphenyl is found to be constant with wavelength of exciting radiation but apparent variations are found for *trans*-stilbene and particularly for anthracene crystals. These variations are due to surface escape of fluorescence the extent of which is sensitive to the depth of penetration of the exciting light in crystals which re-absorb their molecular fluorescence. The fluorescence excitation spectrum thus provides information about the absorption spectrum of the crystal. For excitation of the 001 crystal face of anthracene absorption maxima are found at 3960 Å, 3740 Å, 3540 Å, 3370 Å, 3220 Å; these wavelengths are rather longer than hitherto reported. The first fluorescence maximum lies at 3980 Å hence the 0–0 absorption and fluorescence transitions effectively coincide.

Measurements made with an anthracene crystal having a very thin surface quenching layer of low fluorescence efficiency show that migration of electronic excitation energy through the crystal lattice occurs initially by a short range and non-radiative process. Possible transfer mechanisms are discussed and it is concluded that the most probable process is a quantum-mechanical resonance exchange between adjacent molecules resulting in a random diffusion of the excitation energy through the lattice. This is followed by the secondary and trivial process of fluorescence emission and re-absorption.

The absolute photofluorescence quantum efficiencies of *p*-terphenyl and *trans*-stilbene crystals are measured by comparison with anthracene for which the quantum efficiency is 0·80 ± 0·05. At a temperature of 290°K the efficiencies of *p*-terphenyl and *trans*-stilbene crystals are respectively 0·52 and 0·65.

§ 1. INTRODUCTION

MANY fluorescent organic crystals are opaque to part of their own fluorescence emission. As a consequence the fluorescence of the crystal often differs considerably from that of the molecule. The most obvious modification of the molecular fluorescence is suppression of a part since only those wavelengths to which the crystal is transparent are observed. This has been studied quantitatively for some crystals by Wright (1954 a) and Birks and Wright (1954) and the existence of considerable overlap of absorption and fluorescence spectra verified in certain cases. In addition to this change of the spectrum the decay time of the fluorescence from the crystal is increased (Birks and Little 1953) and its quantum efficiency is reduced (Wright 1955 a) by the cascade process of re-absorption and re-emission of photons in the region of spectral overlap.

---

† Now at Mullard Research Laboratories, Salfords, Redhill, Surrey.

A further result of self-absorption is the enhancement of surface escape of fluorescence (Wright 1955 a). Because of the high extinction coefficients of the crystals the exciting light is rapidly absorbed to form a surface layer of excited molecules. During the initial emission from this layer a certain proportion of fluorescence, determined by the optical constants of the crystal, will escape immediately through the surface. The remainder travels into the crystal, a proportion finding it transparent and escaping, the rest being re-absorbed to produce excited molecules lying rather deeper within the crystal than the initial excited layer. This cascade process of re-absorption, re-emission and surface escape continues until no more excited molecules remain in the crystal. Evidently the total proportion of fluorescence which escapes through the crystal surface is influenced by the extent of self-absorption of fluorescence in the crystal. It is also influenced by the depth of penetration of the exciting light. When the initially excited molecules are produced deep within the crystal a considerable part of the fluorescence which would normally escape through the surface is absorbed before it reaches the surface. Consequently the fluorescence intensity escaping through the crystal surface is less than for wavelengths for which the extinction coefficient is high and which are absorbed in the surface layers of the crystal. Experiments have been made to verify this for crystals of *p*-terphenyl, *trans*-stilbene and anthracene and are reported in this paper; preliminary measurements for anthracene have previously been reported (Wright 1955 b). In addition to these measurements the relative quantum efficiencies of these crystals are estimated and converted into absolute efficiencies by comparison with the recently determined figure for anthracene (Wright 1955 a).

§ 2. EXPERIMENTAL METHOD

The experimental method consisted of illuminating the top surface of the crystal with monochromatic exciting light and measuring the intensity of fluorescence transmitted through the base of the crystal. This observation was repeated for exciting wavelengths throughout the whole absorption region of the crystal above 2200 Å, the fluorescence intensity at each exciting wavelength being normalized to the same quantum flux of incident light. In this way the fluorescence excitation spectrum of each crystal was measured.

The crystals used for these measurements were 1 cm cubes as supplied by the Larco Nuclear Instrument Company for scintillation counting. The crystal under investigation was placed on a thin Perspex disc the underside of which was shaped to fit the cathode of an R.C.A. type 5819 photomultiplier tube mounted vertically as illustrated in figure 1. Optical contact between surfaces was improved

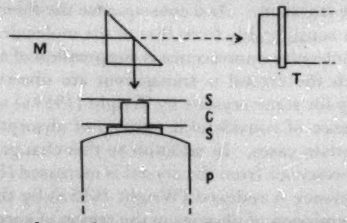


Figure 1. Arrangement of crystal mounting. M, exit beam from Beckman monochromator; T, thermopile or quartz photocell; S, screen; C, crystal; P, photomultiplier.

by the introduction of thin films of glycerine (refractive index = 1·477). Crystal faces were polished gently on soft tissue paper before mounting. The crystal surface exposed to the exciting light was limited to a circular area 2 mm in diameter with a black paper screen; a second screen with a 1 cm square aperture covered the Perspex disc to ensure that the photomultiplier received only fluorescence transmitted through the base of the crystal. Fluorescence was excited by illuminating the top surface of the crystal with monochromatic light from a Beckman Model D.U. spectrophotometer. For fluorescence measurements the horizontal exit beam of this instrument was deflected vertically downwards with a quartz prism.

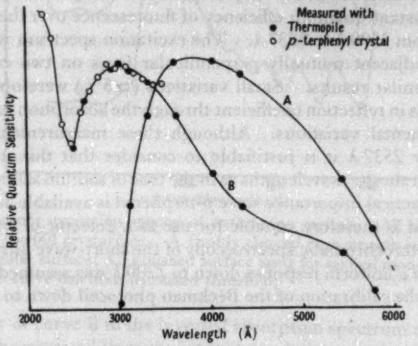


Figure 2. Spectral sensitivity curves. A : R.C.A. type 5819 photomultiplier; B : Beckman quartz photocell.

In order to normalize the observed fluorescence intensities to the same quantum flux of exciting light it was necessary to measure the relative quantum intensity of the monochromator exit beam. This was measured with the quartz photocell of the Beckman instrument which had been previously calibrated for this purpose. To make this calibration the hydrogen lamp of the spectrophotometer had been replaced with a 200-watt mercury arc lamp and the relative sensitivity of the photocell measured with a Hilger type F.T.16 linear vacuum thermopile for eleven strong lines in the wavelength region 5780 Å to 2537 Å. The spectral sensitivity curve obtained is shown in figure 2; the extension down to a wavelength of 2200 Å was made using a *p*-terphenyl crystal as described later. The spectral sensitivity curve of the 5819 photomultiplier was measured at the same time and is also shown in this diagram.

§ 3. RESULTS AND DISCUSSION

The description and discussion of the experimental results may be conveniently divided into two sections dealing respectively with measurement of the fluorescence excitation spectra and measurement of the quantum efficiencies of fluorescence.

3.1. *Fluorescence Excitation Spectra*

p-terphenyl.

Spectral measurements which have been made for this crystal show that self-absorption is negligible. Consequently the excitation spectrum will measure the relative quantum efficiency of the molecular fluorescence when excited with different wavelengths. Many measurements of this kind have been made for organic molecules in solution and have invariably shown the quantum efficiency to be constant, but measurements for crystals are scarce. An exception is sodium salicylate which has been investigated by several experimenters (e.g. Watanabe and Inn 1953) and shown to have a constant quantum efficiency for excitation wavelengths down to 850 Å. It is of interest that there appear to be no significant changes in quantum efficiency for excitation with wavelengths in the neighbourhood of the molecular ionization potential.

The present measurements for *p*-terphenyl show that this molecule too has effectively a constant quantum efficiency of fluorescence over the measured wavelength range from 3500 Å to 2537 Å. The excitation spectrum was measured for each of three adjacent mutually perpendicular faces on two crystals; the two crystals gave similar results. Small variations ($\simeq 5\%$) were observed and were due to variations in reflection coefficient through the absorption band of the crystal and to experimental variations. Although these measurements could not be extended below 2537 Å it is justifiable to consider that this uniform response extends to much shorter wavelengths as in the case of sodium salicylate. This is of considerable practical importance since *p*-terphenyl is available as large crystals of high quality and is therefore suitable for use as a detector of constant quantum efficiency for heterochromatic spectroscopy of the short-wave ultra-violet. In the present instance a uniform response down to 2200 Å was assumed and the crystals used to extend the calibration of the Beckman photocell down to this wavelength.

trans-stilbene.

Fluorescence excitation spectra for this material were measured for a smooth freshly cleaved 001 crystal face and for three adjacent mutually perpendicular faces on a 1 cm scintillation crystal.

For the freshly cleaved face intensity variations of about 10% were observed. Since spectral measurements which have been made for this crystal show that it absorbs approximately half of the molecular fluorescence, these variations were interpreted as being due to the different penetration depths of the exciting light at different wavelengths affecting the proportion of fluorescence escaping through the surface. This interpretation indicated that, for the crystal as for the solution, there is little vibrational structure in the absorption band. For the polished surfaces of the scintillation crystal intensity variations of about 30% were observed. These surfaces, however, were not freshly cleaved and it is possible that some chemical deterioration had occurred. In addition, these surfaces were less smooth than the cleaved surface, consequently surface escape effects were enhanced.

Anthracene.

Crystals of anthracene re-absorb at least three-quarters of the molecular fluorescence. They are, therefore, very suitable for investigating the influence of self-absorption upon surface escape of fluorescence. Some excitation spectra measured for crystalline anthracene are shown in figure 3. These all correspond to excitation of the 001 crystal face. Ordinate scales are adjusted so that the curves coincide for the longest excitation wavelengths used. Curve A was given by a freshly cleaved glass smooth surface; curve B was given by a freshly cleaved surface after gentle scraping with a clean blade to produce a translucent surface. These two curves illustrate very well the enhancement of surface escape of fluorescence with an irregular surface. Curve C was given by the smooth surface of a 1 cm cube scintillation crystal and is discussed in detail later.

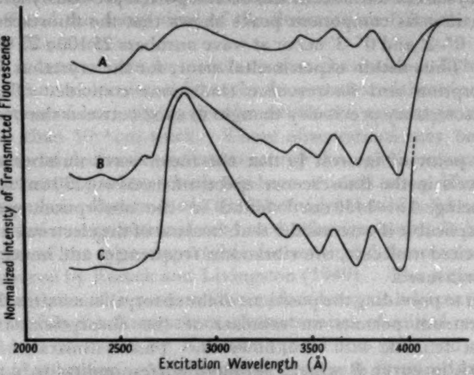


Figure 3. Fluorescence excitation spectra of anthracene crystals. All the curves are for excitation of the 001 crystal face. A, freshly cleaved glass smooth surface; B, freshly cleaved irregular surface; C, polished surface aged for several months; the ordinate scale for this curve has been increased threefold.

The similarity of curve B to the inverted absorption spectrum of anthracene is evident and is to be expected since the intensity variations are due to the variations in the extinction coefficient of the crystal. Since the minima of the excitation spectrum correspond to absorption maxima this curve provides much useful information about the absorption spectrum of the crystal. This is extremely difficult to measure directly because of the high extinction coefficients involved. Previous measurements have been made by Seshan (1936) who gives the positions of the absorption maxima and Kortum and Finckh (1942) who measure actual extinction coefficients.

For the first absorption region the absorption maxima correspond to the electronic transitions $0'-0''$; $0'-1''$; $0'-2''$; $0'-3''$, and $0'-4''$. Using the fluorescence excitation spectrum these occur at wave numbers $\nu = 25\,200$; $26\,700$; $28\,200$; $29\,700$, and $31\,000\ \mathrm{cm}^{-1}$. The mean wave number spacing of the vibrational levels is $\Delta\bar{\nu} = 1430\ \mathrm{cm}^{-1}$ which is the accepted value.

These measurements show that the absorption maxima of the crystal lie at rather longer wavelengths than those reported by Seshan (1936) and Kortum and Finckh (1942). This is possibly because these authors were necessarily obliged to use very thin crystalline films for direct transmission measurements whereas the present measurements were made on large single crystals and are therefore more indicative of the true absorption spectrum of the crystal. These differences may be illustrated with reference to the first absorption maximum. For solution in dioxane this lies at a wavelength 3770 Å; for the crystal Kortum and Finckh report a wavelength 3910 Å; the present measurements give a wavelength 3960 Å. According to the measurements of Kortum and Finckh the mean wave number difference between the absorption spectrum for solution in dioxane and for the crystal is $\Delta\bar{\nu} = 900\ \mathrm{cm}^{-1}$; the present measurements give $\Delta\bar{\nu} = 1200\ \mathrm{cm}^{-1}$.

An interesting result of these measurements is that the 0′–0″ absorption transition coincides with the 0″–0′ fluorescence transition. The molecular fluorescence spectrum of crystalline anthracene has been reported previously and an analysis of this spectrum into its component peaks shows that the fluorescence transitions 0″–0′; 0″–1′; 0″–2′ and 0″–3′ occur at wave numbers 25 100; 23 900; 22 600 and 21 400 cm^{-1}. Thus within experimental error, for the crystal as for the vapour, the 0–0 absorption and fluorescence transitions coincide. This observation reduces the discrepancy previously thought to exist between the positions of these transitions.

A further point of interest is that the mean wave number spacing of the vibrational levels in the fluorescence spectrum is $\Delta\bar{\nu} = 1230\,\mathrm{cm}^{-1}$. This is less than the spacing $\Delta\bar{\nu} = 1430\,\mathrm{cm}^{-1}$ found in the absorption spectrum. This difference is explicable if we consider that, because of the electronically compressed state of the excited molecule, the vibrational frequencies and hence the spacing of the levels are increased.

In addition to providing the positions of the absorption maxima the fluorescence excitation spectrum permits an estimate of the absorption strength. Such estimates must be made with care, however. This is illustrated by the minima at about 2500 Å in curve B which, although corresponding to a much stronger absorption peak, lie little lower than that at 3960 Å. The reason is that at each of these wavelengths the initial excited molecules are produced in a surface layer of negligible thickness compared with the mean path lengths for re-absorption of fluorescence. Under these circumstances variations in the thickness of the initial excited layer do not appreciably affect the intensity of the transmitted fluorescence.

The fluorescence excitation spectrum illustrated by curve C of figure 3 is of interest in that it was given by the smooth 001 face of a 1 cm cube scintillation crystal which had been kept mostly in the dark, but exposed to the atmosphere, for several months. It had not previously been used. Although illustrated for only one surface this spectrum is characteristic of the other surfaces of the crystal and of other crystals. The very low efficiency of this fluorescence, which for some wavelengths is less than a tenth of that from a freshly cleaved surface, can only be attributed to chemical deterioration (probably oxidation resulting in the formation of the photo-oxide and anthraquinone) producing in the surface molecular layers impurity molecules capable of quenching the anthracene fluorescence. The thickness of the quenching layer can be estimated as less than 10^{-4} cm since for an excitation wavelength of 2850 Å, which has a mean penetration depth of this order, the fluorescence efficiency has risen appreciably showing that this wavelength penetrates well past the quenching layer.

If the impurity molecules were present in sufficient concentration to absorb the major part of the exciting light and account for the low fluorescence efficiency in this way the fluorescence excitation spectrum would correspond to the absorption spectrum of the impurity molecule. The maxima and minima of this occur in exactly the same positions as for pure anthracene, however. This strongly suggests that the exciting light is still absorbed by anthracene molecules but that the fluorescence of these is quenched by a relatively small concentration of impurity molecules. On this explanation one would expect the additional quenching process to produce a major decrease in the decay time of the fluorescence. Such a decrease would be masked to a large extent by the cascade process of re-absorption and re-emission of photons in the crystal but one would still expect the observed decay time to be appreciably less than for excitation of a fresh surface. This was found experimentally. The decay time of the fluorescence was measured (assuming an exponential decay) to be 12 mμ sec; the damaged surface layer was then cleaved off and the decay time of the fluorescence from the fresh surface measured to be 22 mμ sec.

From these measurements we conclude that the incident light is absorbed mainly by anthracene molecules and that the fluorescence of these is quenched by a small concentration of impurity molecules which are present in a surface layer appreciably less than 10^{-4} cm thick. These observations may be explained in terms of migration of electronic excitation energy through the crystal lattice to the quenching molecule, a process which has been studied using dilute solutions of fluorescent molecules. Bowen *et al.* (1949) consider that in solid crystalline solutions the excitation energy is transferred to the foreign molecule by a resonance exchange process of the type responsible for sensitized fluorescence in gases, a mechanism favoured by Franck and Livingston (1949). The author (1953 a) has previously shown that such a process taking place between adjacent molecules can quantitatively explain the fluorescence characteristics of mixed organic crystals. This mechanism is also used by Kallmann and Furst (1950) to explain energy transfer in liquid solutions. An alternative process is favoured by Birks (1953) who considers that for both solid and liquid systems the excitation energy is transferred solely by fluorescence emission and re-absorption. In most cases both of these proposed mechanisms lead to similar experimental features of fluorescence and it is difficult to decide which, if either, is responsible. The present measurements, however, differentiate between these mechanisms.

Unless quenching of most of the initially excited molecules takes place before these can emit fluorescence at least a quarter will emit fluorescence photons to which the crystal is transparent and which will therefore escape from the crystal. Even if all the re-absorbed photons were captured by impurity molecules and quenched, which is by no means likely, this would still result in the fluorescence intensity being several times greater than that observed. Since, then, quenching occurs before the molecules emit, it must be a non-radiative process. The thickness of the quenching layers is less than 10^{-4} cm and hence the operative process must be a short range mechanism. This is the case for the resonance interaction process mentioned above which can occur only over molecular distances. For the radiative process mean path lengths for re-absorption of the greater part of the fluorescence are ten to a thousand times greater than this. We conclude then that the present measurements provide direct evidence for a non-radiative short-range mechanism for transfer of the electronic excitation energy.

There appear to be three possible types of process which satisfy these conditions. They are electron migration, exciton migration or a quantum-mechanical exchange process between molecules; conditions in the organic crystal lattice exclude the first two of these but are very suitable for the occurrence of the third. Except in insulating crystals possessing a conductivity band, electron migration is only possible if ionization occurs; this will take place during the scintillation process but is clearly excluded for photofluorescence. Exciton migration is a process, possible in certain types of insulating crystals, in which an electron in the filled band, acquiring insufficient energy to raise it into the conduction band, becomes trapped in the field of its own positive hole. The combination may move through the crystal lattice transporting excitation energy but not charge. Again this process is excluded in organic crystals. The term exciton transfer has also been used (Frenkel 1931) to describe an extremely rapid movement of excitation energy through the component parts of a strongly coupled system. For the crystals under discussion there are no indications in the absorption spectrum that this necessary strong coupling between molecules exists and we may therefore discard this mechanism. For the remaining transfer mechanism there are two possibilities. The first of these was suggested by Förster (1948) who shows that coupling between molecules due to dipole–dipole interaction can result in transfer of electronic excitation energy. However the types of molecules under consideration have such very weak dipole strengths that this process is effective only over very short distances. These distances are so short that appreciable transfer by this process in these crystals may be ruled out. The other possibility has been mentioned previously and involves a quantum-mechanical resonance exchange between the electron orbitals of the excited molecule and an unexcited molecule resulting in transference of the excitation energy. This process is discussed by Massey and Burhop (1952). Clearly this process is quite possible in organic crystals since the molecules retain their individuality in the crystal lattice yet are sufficiently near for direct interaction to occur between the expanded electron orbital of the excited molecule and those of adjacent molecules.

Accepting this latter transfer mechanism the excitation energy migrates in a random fashion through the crystal lattice by repeated transfer between adjacent molecules until it is eventually emitted as fluorescence or dissipated non-radiatively by a molecule of the host lattice or is captured by a foreign molecule. If the non-radiative transition probability of the foreign molecule is high the excitation energy will be dissipated before it is transferred again or if the electron levels of this molecule are below those of the host molecules the excitation energy will remain with the foreign molecule until dissipated radiatively or non-radiatively. It has been shown previously (Wright 1952, 1953 a) that this mechanism of energy transfer can quantitatively explain the fluorescence characteristics of mixed organic crystals if the mean transfer frequency of the excitation energy is similar to the thermal vibration frequency of the lattice. In the pure crystal lattice this random diffusion of the excitation energy will be a fundamental part of the photo-fluorescence and scintillation processes. The excitation energy will be transferred over considerable distances and fluorescence emission will, in general, occur at some distance from the point of absorption. (An approximate calculation indicates that mean distances of transfer in typical fluorescent crystals range from 10–100 molecular diameters.) Energy migration is of particular importance in the scintillation process since it will result in a rapid expansion of the primary excitation column formed along the track of the absorbed particle. This will influence the nature of the quenching mechanisms operating in the scintillation column (Birks 1953, Black 1953, Wright 1953 b, 1954 b) and also the rate of emission of fluorescence during the scintillation.

3.2. *Absolute Quantum Efficiencies of Fluorescence*

These measurements were made using wavelengths just within the transmission edge of the crystal. These penetrate deeply into the crystal and surface effects, as described in the previous section, are thereby eliminated.

When the incident light is absorbed in the crystal it is converted into fluorescence with a quantum efficiency less than unity. The fluorescence of the crystal is emitted in all directions and only a part of this is received by the photocathode. Consequently the multiplier output current, when corrected for the different spectral sensitivity of the photocathode to the fluorescence light and the exciting light, is less than when the exciting light is received directly by the photocathode. The ratio of the corrected multiplier currents gives the product of the quantum efficiency q and the efficiency of collection of fluorescence f.

Using the same experimental procedure as described previously this ratio was measured for a range of wavelengths in the neighbourhood of the transmission edge for a number of crystals. A typical set of uncorrected measurements for *p*-terphenyl, *trans*-stilbene and anthracene are illustrated in figure 4. For each crystal measurements were made for excitation of each of three adjacent mutually perpendicular faces. These curves are all very similar and may be discussed with particular reference to those for anthracene.

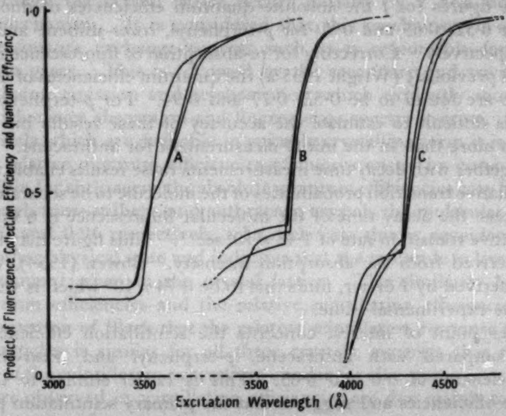


Figure 4. Illustrating the dependence of the product of fluorescence collection efficiency and quantum efficiency upon the exciting wavelength. A, *p*-terphenyl; B, *trans*-stilbene; C, anthracene.

The transmission edge of this crystal is at 4270 Å and at wavelengths immediately below this the crystal fluoresces with constant quantum efficiency. At slightly shorter wavelengths the intensity of the transmitted fluorescence decreases slightly because of a gradual decrease in the collection efficiency of fluorescence as the penetration depth of the exciting light decreases. For exciting wavelengths shorter than about 4000 Å the penetration depth is small, photofluorescence is essentially a surface phenomenon and the fluorescence intensity becomes very small due to surface quenching of fluorescence as described in the previous section.

Since the observed fluorescence intensity varies only slightly for different crystal orientations the mean value of fq for all three curves may be taken. For *p*-terphenyl, *trans*-stilbene and anthracene we find $\bar{f}q = 0{\cdot}24$, 0·39, 0·38 at excitation wavelengths 3460 Å, 3650 Å and 4270 Å respectively. These wavelengths are the respective transmission limits of the 1 cm thick crystal. Assuming isotropic optical properties and assuming total internal reflection for angles of incidence greater than the critical angle we may calculate the collection efficiency for these three crystals to be 0·37, 0·47 and 0·37 respectively. For the particular case of anthracene the quantum efficiency has been measured to be 0·80 ± 0·05 hence we may measure the collection efficiency for this crystal to be 0·47. That this is different from the calculated value is to be expected since the crystals are by no means optically isotropic nor are their surfaces optically smooth. However, since each of these hydrocarbons crystallizes in the monoclinic system their optical properties are similar; it is therefore justifiable to consider that for each of these crystals, since the collection geometry is the same in each case, there will be the same relative difference between the calculated and the actual collection efficiency. Thus, normalizing to a value of 0·47 for anthracene we find collection efficiencies of 0·47 and 0·59 for *p*-terphenyl and *trans*-stilbene respectively. Using these figures for $\bar{f}$ the absolute quantum efficiencies of fluorescence are found to be 0·52, 0·65 and 0·80 for *p*-terphenyl, *trans*-stilbene and anthracene crystals, respectively. Correcting for re-absorption of fluorescence in the crystal as described previously (Wright 1955 a) the quantum efficiencies of the molecular fluorescence are found to be 0·52, 0·77 and 0·94. For *p*-terphenyl and *trans*-stilbene it is difficult to estimate the accuracy of these results but the error is probably no more than in the initial measurement for anthracene.

Used together with decay time measurements these results enable the radiative and non-radiative transition probabilities of the molecule to be separately obtained. For anthracene the decay time of the molecular fluorescence is 6·4 mμsec which gives a radiative transition rate of $1{\cdot}43 \times 10^8$ sec^{-1}. This figure may be compared with that derived from the absorption intensity. Bowen (1954), who uses an expression derived by Förster, finds this to be $0{\cdot}74 \times 10^8$ which is within a factor of two of the experimental value.

A further point of interest concerns the scintillation efficiencies of these crystals. Compared with anthracene, *p*-terphenyl and *trans*-stilbene have relative efficiencies of 0·6 and 0·65. This is rather similar to their relative fluorescence efficiencies and suggests that the primary scintillation processes for these crystals are similar and that the differences observed in scintillation efficiencies are due mainly to differences in photofluorescence characteristics. This is also suggested by the work of Black (1953) who shows that the relative responses of these crystals to electrons and alpha particles are similar.

§ 4. Conclusions

The experimental observations reported show that for crystals which strongly re-absorb their own fluorescence there is an apparent variation of quantum efficiency with exciting wavelength. For *p*-terphenyl which shows negligible self-absorption this effect is absent. Consequently this crystal is very suitable as a detector of constant quantum efficiency for heterochromatic spectroscopy in the short-wave ultra-violet. For *trans*-stilbene and particularly for anthracene, however, appreciable variations occur in the intensity of the transmitted fluorescence especially if the surface is not smooth. These variations provide much useful information about the absorption structure of the crystal. Such measurements may be extended into the vacuum ultra-violet and soft x-ray regions of the spectrum and may provide valuable information about higher electronic excited states and ionization potentials of the molecule in the crystalline state.

Surface escape of fluorescence increases as the depth of penetration of the excitant decreases. This produces an apparent decrease in the scintillation efficiency of the crystal over and above any decrease due to quenching mechanisms in the primary excitation column. This will result in a lower scintillation response for short-range particles if incident externally compared with the response if produced internally. This is found experimentally by Birks and Brooks (1954) and Brooks (1955) (although this is disputed by Robinson and Jentschke 1954) and has been discussed for anthracene (Wright 1955 c). Such effects are particularly appreciable if the surface is not fresh and smooth. For *p*-terphenyl these effects are very small and, notwithstanding its lower efficiency, this material would appear more suitable for use in scintillation counters.

The very low fluorescence efficiency shown for surface excitation of an anthracene crystal with a very thin surface quenching layer shows that energy migration through the crystal lattice occurs initially by a non-radiative and short-range mechanism. It is considered that this mechanism is a quantum-mechanical resonance exchange process such as is responsible for sensitized fluorescence in gases. This is followed by the secondary and trivial radiative process of photon emission and re-absorption which evidently must occur to some extent whenever absorption and fluorescence spectra overlap.

For materials which are available in suitable crystalline form it is possible to measure the relative quantum efficiencies of fluorescence; by comparison with the known value for anthracene the absolute quantum efficiencies may be obtained. For *p*-terphenyl, *trans*-stilbene and anthracene crystals the molecular efficiencies are 0·52, 0·77 and 0·94 respectively. In each case this is considerably higher than in any other physical state and indicates that the molecule is least perturbed by its environment when in the crystal lattice. The similarity between the relative quantum efficiencies and the relative scintillation efficiencies, together with the observation of Black that the relative scintillation response to electrons and alpha particles is similar for all three crystals, suggests that the primary mechanisms of the scintillation process are similar for all organic crystals and that the differences observed in scintillation efficiencies are due mainly to differences in photofluorescence characteristics.

ACKNOWLEDGMENTS

The author wishes to thank Mr. T. D. S. Hamilton for making the decay time measurements quoted and the South African Council for Scientific and Industrial Research for a grant to purchase research equipment.

REFERENCES

BIRKS, J. B., 1953, *Scintillation Counters* (London : Pergammon Press).
BIRKS, J. B., and BROOKS, F. D., 1954, *Phys. Rev.*, **94**, 1800.
BIRKS, J. B., and LITTLE, W. A., 1953, *Proc. Phys. Soc.* A, **66**, 921.
BIRKS, J. B., and WRIGHT, G. T., 1954, *Proc. Phys. Soc.* B, **67**, 657.
BLACK, F. A., 1953, *Phil. Mag.*, **44**, 263.
BOWEN, E. J., 1954, *Trans. Faraday Soc.*, **50**, 97.
BOWEN, E. J., MIKIEWICZ, E., and SMITH, F. W., 1949, *Proc. Phys. Soc.* B, **62**, 26.
BROOKS, F. D., 1955, *Thesis*, Rhodes University.
FÖRSTER, T., 1948, *Ann. Phys., Lpz.*, **2**, 55.
FRANCK, J., and LIVINGSTON, R., 1949, *Rev. Mod. Phys.*, **21**, 505.
FRENKEL, J., 1931, *Phys. Rev.*, **37**, 1276.

Kallmann, H., and Furst, M., 1950, *Phys. Rev.*, **79**, 857.
Kortum, G., and Finckh, B., 1942, *Z. Phys. Chem.* B, **52**, 263.
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: Clarendon Press).
Robinson, W. H., and Jentschke, W., 1954, *Phys. Rev.*, **95**, 1412.
Seshan, P. K., 1936, *Proc. Ind. Acad. Sci.*, **3a**, 148.
Watanabe, K., and Inn, E. C. Y., 1953, *J. Opt. Soc. Amer.*, **43**, 32.
Wright, G. T., 1952, *Thesis*, University of Birmingham; 1953 a, *Proc. Phys. Soc.* A, **66**, 777; 1953 b, *Phys. Rev.*, **91**, 1282; 1954 a, *South African Congress of Physics, Pretoria*, July 1954; 1954 b, *Phys. Rev.*, **96**, 569; 1955 a, *Proc. Phys. Soc.* B, **68**, 241; 1955 b, *Phys. Rev.*, in the press; 1955 c, *Ibid.*, in the press.

Ferromagnetic Resonance in Nickel and in some of its Alloys

BY K. J. STANDLEY AND K. H. REICH†
Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 29th April 1955 and in final form 19th May 1955

Abstract. From microwave resonance measurements, g values and line widths for nickel and for Ni–Cu, Ni–Al, Ni–Sb and Ni–Mn alloys with high nickel content have been determined. Measurements were made at a wavelength of 1·22 cm and usually in the temperature range 20–200°c.

A constant temperature independent g value of $2{\cdot}19 \pm 0{\cdot}02$ was found for all specimens except the Ni–Mn alloys; for these a decrease in g value with increase in manganese content occurred, reaching 2·11 for the 13·5 atomic per cent Mn alloy. An explanation of this behaviour is given.

Absolute absorption measurements were made and the line widths are expressed in terms of the damping constant $1/T_2$ introduced by Bloembergen (1950). For a given alloy the shape of the curve of $1/T_2$ against reduced temperature is similar to that found by Bloembergen for nickel but varies from alloy to alloy. A single curve of exponential form is found to represent the results when $1/T_2$ is plotted against saturation intensity of magnetization M, an increase in $1/T_2$ accompanying a decrease in M. A possible explanation is qualitatively presented.

---

§ 1. INTRODUCTION

MICROWAVE resonance absorption has been studied in nickel and in alloys of this metal with copper, aluminium, antimony and manganese. The object of the investigation was to examine the variation of the spectroscopic splitting factor g and the damping constant $1/T_2$ of the ferromagnetic nickel in various degrees of dilution with non-ferromagnetic metals. The measurements were carried out on polycrystalline specimens at a wavelength of 1·22 cm and in the temperature range from 20°c to 200°c or to the Curie point, whichever was the lower.

The high-frequency power loss in the material at constant frequency ν was measured as a function of a steady magnetic field H_z which was applied perpendicular to the high-frequency field H_x. Under these conditions the loss rises to a maximum at a field $H_z^{\max}$, and Kittel (1948) and others have shown that the resonance condition for polycrystalline material is

$$\nu = \frac{ge}{4\pi mc}\{(H_z^{\max} + (N_y - N_z)M)(H_z^{\max} + (N_x - N_z)M)\}^{1/2}. \qquad \ldots\ldots(1)$$

Here M is the intensity of magnetization in the material, assumed to be saturated in the field $H_z^{\max}$; N_x, N_y and N_z are the demagnetization factors of the specimen in the three coordinate directions, and the other symbols have their usual meanings.

The relaxation processes occurring in ferromagnetic resonance are not clearly understood, and the shape and width of the resonance curves have not yet received satisfactory theoretical explanation. Several empirical damping terms in the classical equations of motion have been suggested (Kittel 1947, 1948, Bloembergen 1950, Yager *et al.* 1950); by suitable choice of the damping constant in such terms, a reasonable agreement between theoretical and experimental curves may be obtained. The present investigation was concerned mainly with a *comparison* of the damping constants of materials under similar experimental conditions, and it was found convenient to use the equations given by Bloembergen (1950) and based upon Bloch's nuclear resonance equations.

---

† Present address 22A, Rheydt, Nordstrasse 172, W. Germany.

The relevant equations, referred to unit volume of the material, are

$$dM_{x,y}/dt=\gamma[\mathbf{M}\times\mathbf{H}]_{x,y}-M_{x,y}/T_2$$
$$dM_z/dt=\gamma[\mathbf{M}\times\mathbf{H}]_z-(M_z-M)/T_1$$

where $\mathbf{H}$ is the effective field within the specimen. Provided that the damping constant $1/T_2\ll\nu$, the resonance equation (1) is unaltered at the low microwave powers employed, and

$$1/T_2=2\gamma^2M[H_z^{\max}+(N_y-N_z)M]/\nu\mu_R^{\max}. \qquad \ldots\ldots(2)$$

In this equation, $\gamma=ge/2mc$ and, writing the complex permeability in the form $\mu=\mu_1-i\mu_2$, $\mu_R=(\mu_1^2+\mu_2^2)^{1/2}+\mu_2$; $\mu_R^{\max}$ is the value of μ_R found in the field $H_z^{\max}$.

Thus the determination of values of g and $1/T_2$ necessitates the measurement of $H_z^{\max}$ and $\mu_R^{\max}$; in addition, a knowledge of the frequency, intensity of magnetization and demagnetization factors is required.

The method and results are given in detail in §§2 and 3 respectively, and are discussed in §4 of this paper.

§ 2. Experimental Apparatus and Procedure

2.1. *General Arrangement of the Apparatus*

The general form of the apparatus (figure 1) has already been described (Adam and Standley 1953). The cavity was tuned by a micrometer driven non-contact plunger, and the specimen formed the bottom wall of the cavity, which resonated in an H_{112} mode, heat insulation being provided by water jackets surrounding the input and output waveguides and the micrometer block.

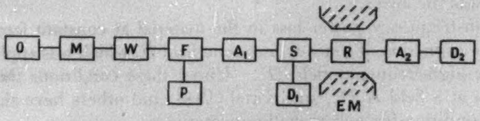


Figure 1. Block diagram of the apparatus.

O, klystron oscillator; M, matching unit; W, wavemeter; F, directive feed; P, power monitor; A, attenuator; S, standing wave detector; D, detector; R, cavity resonator; EM, electromagnet.

2.2. *Experimental Method*

It was shown previously (Adam and Standley 1953) that Q_0, the unloaded Q of the cavity, may be written in the form

$$1/Q_0=a+b(\mu_R\rho)^{1/2} \qquad \ldots\ldots(3)$$

where a and b are constants for the cavity at a given temperature and ρ is the electrical resistivity of the specimen. In this investigation absolute values of μ_R were required so that a determination of a and b was necessary.

The loaded Q of the cavity, Q_L, was determined at constant frequency by measuring the detuning of the cavity required to reduce the power transmitted to a matched load to half its maximum value. For the transmission type cavity used, $1/Q_L=(1+\beta_1+\beta_2)/Q_0$ where β_1 and β_2 are the coupling parameters of the input and output guides. The diameters of the circular irises coupling the cavity to the input and output waveguides were adjusted until the measured values of β_1 and β_2 were identical, and thus $1/Q_L=(1+2\beta)/Q_0$. The parameter β was found from the standing wave ratio r set up in the input waveguide, and in the present case of an undercoupled cavity was equal to $1/r$. The measured standing wave ratio r' was corrected for the finite detector probe impedance, for the power loss in the guide walls between the probe and the iris, and for the loading of the cavity by the output guide. Measurements of Q_L and β thus enabled Q_0 to be found. The theory underlying these measurements and details of the various corrections may be found, for example, in Montgomery (1947).

The constants a and b of the cavity were found from values of Q_0 determined when the ferromagnetic specimen was replaced by highly polished, strain free non-ferromagnetic materials whose resistivities were known at the temperatures used. In general, silver and an alloy similar to manganin were used, but other metals provided additional checks of the calibration.

The required data for the ferromagnetic specimens could have been obtained from measurements of Q_L and β in a series of applied fields H_z. A shorter and more convenient procedure, however, was that outlined below.

Q_L was first determined in the remanent field of the electromagnet. With the cavity tuned to resonance, a measure of the transmitted power P, proportional to Q_L^2, was given by the rectified output current of a silicon tungsten crystal detector (D_2, figure 1). This current was measured by the deflection θ_0 of a sensitive low resistance galvanometer. The magnetic field was then increased and the deflection θ recorded for a series of known fields up to 13850 oersteds. Since the crystal was found experimentally to be working as a square law detector

$$Q_L^2 \propto P \propto \theta.$$

To obtain values of Q_0 at each field strength without measuring r' use was made of the equation

$$1/Q_0=1/Q_L-2/Q_E \qquad \ldots\ldots(4)$$

where Q_E is the external or radiation Q of the cavity. It is defined as Q_0/β and was determined from initial measurements with non-ferromagnetic specimens. On the assumption that the resonance absorption in the ferromagnetic material affects only the effective resistance of the cavity and does not alter appreciably the effective reactance, Q_E is a constant independent of the applied field H_z; this was checked experimentally.

The use of equations (4) and (3) thus gave values of $(\mu_R\rho)^{1/2}$. Calculation of μ_R could be effected either by assuming the d.c. resistivity of the material to be unaltered at microwave frequencies (Benson 1953) or by extrapolating the $((\mu_R\rho)^{1/2}, H_z)$ curve to very high field strengths where $\mu_R=1$. An additional check arose from the fact that in the lowest fields μ_R was very close to, but slightly less than, unity.

While the saturation intensities of magnetization of most of the alloys at room temperature were obtained from the literature (e.g. Hoselitz 1952), check measurements were made on most specimens using an oscillation magnetometer of the type described by Griffiths and MacDonald (1951). The 'anti-resonance' in the (μ_R, H_z) curve also enabled M to be found at temperatures well removed from the Curie point (Bloembergen 1950). The demagnetization factors were calculated from published tables and equations (Osborne 1945). Thus, all required parameters in equations (1) and (2) were obtained and values of g and $1/T_2$ could be calculated.

2.3. *Preparation of the Specimens*

Four of the nickel copper alloys were kindly supplied by the Mond Nickel Co. The remaining alloys were prepared by melting together *in vacuo* the requisite amounts of spectroscopically standardized materials, using a high-frequency furnace. The resulting alloy was heated in argon to about 1000°c for 12 hours and slowly cooled to room temperature. After rolling to a sheet about 0·5 mm thick, discs of the required diameter were punched out, polished on fine emery, annealed *in vacuo* and then electrolytically polished. This was usually followed by a final anneal *in vacuo*. Experience showed that this final anneal was often important, removing any small residual surface strains left after polishing; an increase in the maximum power absorption of 5% in some cases was recorded, with a corresponding decrease in line width. Chemical analysis of most of the alloys was carried out using standard volumetric procedures.

2.4. *Experimental Errors*

The main uncertainties in the measured g values arise from the determination of $H_z^{\max}$ and M. The calibration and setting of the magnetic field and the location of the resonance peak are thought to introduce a possible $1\frac{1}{2}\%$ error in $H_z^{\max}$, while M is uncertain to 2%. These, with the smaller errors in the other quantities involved, yield a maximum error of about $2\frac{1}{2}\%$ in g. Certainly it would be unwise to attach much importance to differences in g less than 0·04.

The probable error in $1/T_2$ is more difficult to estimate with certainty. Equation (2) shows the parameters involved in its determination and the several necessary measurements have been discussed above. From a consideration of the uncertainties in these measurements it is thought that absolute values of $1/T_2$ may be in error by as much as 25%. However, *relative* values will be more reliable and changes in $1/T_2$ of more than 12–15% should be significant. Certainly, individual measurements over a period of nearly two years have been reproducible within this limit.

§ 3. EXPERIMENTAL RESULTS

Some of the experimental data are given in tables 1 and 2. In general observations were made at several temperatures but in table 1 are recorded only those data taken at the extreme temperatures and at those intermediate temperatures where significant changes in $1/T_2$ were found. Some comments on individual alloys are given below.

Table 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cu 0·8	347	20	7·9	6050	37·5	4·5	2·18
		200	16·0	5025	27·2	4·8	2·18
Cu 4·6	308	20	12·6	5500	27·8	5·3	2·18
		200	22·5	4260	18·0	6·0	2·21
Cu 9·75	256	20	19·8	4700	20·0	6·1	2·19
		150	28·9	3700	13·7	6·6	2·21
		200	32·5	3100	10·3	7·5	2·22
Cu 14·8	203	20	26·0	3900	14·4	6·7	2·19
		100	32·4	3250	10·5	7·4	2·19
		175	38·3	2100	5·0	9·4	2·22
Cu 24·7	105	20	38·2	2250	5·1	9·8	2·20
		85	44·1	1200	2·2	11·5	2·20
Cu 26·4	89	20	39·8	1830	3·8	10·6	2·18
		75	44·7	1020	1·7	12·4	2·22
Cu 28·5	69	20	41·5	1400	2·6	11·5	2·19
Al 5·1	220	20	20·9	4400	18·3	6·1	2·20
		200	26·3	3150	10·8	6·9	2·20
Al 6·3	190	20	25·0	3750	14·0	6·5	2·19
		100	27·7	3250	10·9	7·1	2·19
		175	29·9	2350	6·5	8·1	2·19
Al 9·5	70	20	35·7	2150	6·0	7·8	2·17
Sb 3·1	250	20	19·5	4250	16·5	6·4	2·17
		150	31·3	3100	10·0	7·2	2·18
		200	35·5	2400	6·3	8·4	2·18
Sb 4·2	190	20	25·0	3600	11·3	7·6	2·17
		150	37·0	2000	5·2	8·8	2·18
		175	41·0	1350	3·2	9·8	2·17
Pure nickel	358	20	6·8	6100	36·1	4·6	2·17
		200	16·0	5150	29·3	4·6	2·18

(1) Alloy (atomic %); (2) Curie temperature (°c); (3) temperature of measurement (°c); (4) resistivity ($\mu\Omega$ cm); (5) $4\pi M$ (e.m.u. cm^{-3}); (6) $\mu_R^{\max}$; (7) $10^9/T_2$ (sec^{-1}); (8) g.

(*a*) *Copper–nickel alloys.*

Seven alloys containing from 0·8 to 28·5 atomic per cent Cu were examined, and the measured g values ranged from 2·17 to 2·22; 24 different measurements were made on this alloy series and each was usually repeated on three different samples. A typical curve of apparent permeability μ_R is given in figure 2.

(*b*) *Aluminium–nickel alloys.*

The alloys used contained 5·1, 6·3 and 9·5 atomic per cent Al. The last alloy is almost at the limit of solubility of aluminium in nickel (Bozorth 1951). Some difficulty was experienced in producing a smooth highly polished surface with these alloys, due possibly to the ready formation of an oxide layer. The nine measured g values fell in the range 2·17 to 2·21.

(*c*) *Antimony–nickel alloys.*

Alloys containing 3·1 and 4·2 atomic per cent Sb were prepared. They were hard and brittle, and good electropolishing was found to be very difficult.

Since it is known that the maximum absorption is very dependent upon the surface condition of the specimen, the recorded values of $1/T_2$ are less reliable than those for the preceding alloys. Slight surface imperfections have little effect upon the resonance field, however, and eight different measurements yielded g values in the range 2·17 to 2·19.

(*d*) *Pure nickel.*

Observations made on several different specimens yielded g values in the range 2·17 to 2·18 at temperatures in the range 20–200°c and the results agree, within experimental error, with those of Bloembergen (1950).

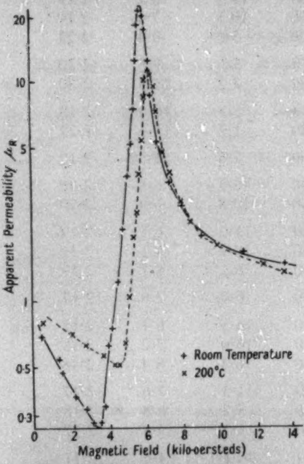


Figure 2. Absorption curves for the alloy containing 9·75 atomic % copper.

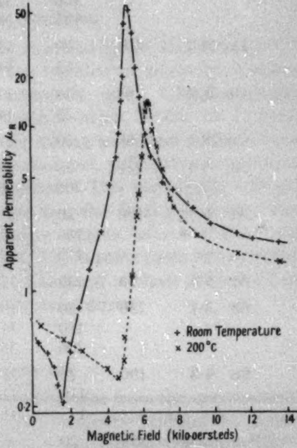


Figure 3. Absorption curves for the alloy containing 10·1 atomic % manganese.

For nickel and the three foregoing alloys, a constant temperature-independent g value was found. The mean value of the 49 different measurements was 2·19 with a mean square deviation of less than 0·003, and all but three of the results were in the range $2{\cdot}19 \pm 0{\cdot}02$.

(*e*) *Manganese–nickel alloys.*

The addition of a small amount of manganese to nickel increases the saturation intensity of magnetization M above that of pure nickel and lowers the Curie point. At about 6 atomic per cent Mn, the increase in M amounts to nearly 10%, thereafter M falls with the addition of further manganese.

Four alloys were prepared containing 1·2, 5·1, 10·1 and 13·5 atomic per cent Mn. The first three alloys lie in the range where the material is generally considered to be disordered (Bozorth 1951), while the last falls close to the border between disordered and ordered states. Observed data are given in table 2 and marked differences from the preceding results are seen. For each of the four alloys, the g value was constant with change in temperature but there is clearly a trend to lower g values with increasing manganese content in the alloy. Certainly the mean value of 2·11 for the 13·5 atomic per cent Mn specimens differs from the value for pure nickel by more than the experimental error. Only in the last two alloys was there a significant change in $1/T_2$ in the temperature range studied. A typical curve of μ_R against H_z for this alloy series is shown in figure 3 and it is seen that the curve shape is essentially the same as for the copper–nickel alloys.

Table 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1·2	340	20	8·8	6250	38·7	4·5	2·17
		200	17·7	5000	30·5	4·3	2·19
5·1	305	20	15·2	6800	42·5	4·5	2·14
		200	25·9	5200	29·7	4·5	2·15
10·1	235	20	26·4	6600	53·6	3·5	2·14
		200	35·4	3650	14·0	5·9	2·13
13·5	190	20	36·0	5800	46·9	3·3	2·12
		150	41·6	3600	13·3	6·2	2·10
		175	43·8	2200	4·5	10·4	2·11

(1) Atomic % Mn; (2) Curie temperature (°c); (3) temperature of measurement (°c); (4) resistivity ($\mu\Omega$ cm); (5) $4\pi M$ (e.m.u. cm^{-3}); (6) $\mu_R^{\max}$; (7) $10^9/T_2$ (sec^{-1}); (8) g.

§ 4. Discussion of Results

4.1. *g values*

An examination of tables 1 and 2 reveals two facts: (i) for pure nickel and for the alloys examined, the g value for a given material remains constant with change of temperature, and (ii) the g values for Ni–Cu, Ni–Al, Ni–Sb alloys are independent of composition and equal to the value for pure nickel (2·19). This is not true for the Ni–Mn alloys where the g value decreases with increase in manganese content, reaching a value of 2·11 for the 13·5 atomic per cent Mn alloy.

Bloembergen (1950) and Bagguley and Harrick (1954) have already reported that the measured g value for nickel is independent of temperature. The application of a simple band model to Ni–Cu alloys does not appear readily to predict a constant g value, but from an atomic standpoint, the observed results follow, in a general way, from the observation that in the first three types of alloy, the alloying material is diamagnetic and may be considered to have the effect of diluting the nickel in the lattice without appreciably altering its magnetic properties. But manganese is paramagnetic, and the observed g value for a Ni–Mn alloy may be considered as a mean of the individual g values for manganese and for nickel.

Consideration of a simple model demonstrates these effects in a semi-quantitative manner. Suppose 1 cm^3 of nickel possesses angular momentum J_{Ni} and magnetic moment M_{Ni}, while J_A and M_A are the corresponding quantities for the alloying material. The undamped equation of motion is $dJ^*/dt = \mathbf{M}^* \times \mathbf{H}$ where J^* and M^* are average values for the alloy. The actual averaging processes are probably not the same for J^* and M^*, but to show the nature of the effect it suffices to average directly and write

$$J^* = (1-x)J_{\mathrm{Ni}} + xJ_A,$$

and

$$M^* = (1-x)M_{\mathrm{Ni}} + xM_A,$$

where x is the fraction of alloying material A in the lattice. Since $M = gJ\beta$, where β is the Bohr magneton, we find

$$g_{\mathrm{eff}} = \frac{(1-x)g_{\mathrm{Ni}}J_{\mathrm{Ni}} + xg_AJ_A}{(1-x)J_{\mathrm{Ni}} + xJ_A}.$$

Clearly, if $J_A=0$, as is the case if A is copper, aluminium or antimony, $g_{\text{eff}}=g_{\text{Ni}}$ as has been found experimentally.

For Heusler alloy the measured g value is 2·01 (Yager and Merritt 1949) and the magnetic moment is close to 4β per Mn atom (Bozorth 1951). Although in that alloy the manganese is in a body centred lattice we may suppose that in the face centred Ni–Mn alloys g_A and J_A differ from g_{Ni} and J_{Ni} and for small concentrations x of manganese we find, approximately,

$$g_{\text{Ni}}-g_{\text{eff}}=(g_{\text{Ni}}-g_A)xJ_A/J_{\text{Ni}}.$$

Neither the experimental accuracy nor the range of x permits the quantitative examination of this relationship, but the trend of the experimental results is, clearly, similar. The averaging processes postulated are almost certainly inadequate but suffice to demonstrate the expected behaviour of the g values.

The mean value of $2{\cdot}19 \pm 0{\cdot}02$ for nickel and certain of its alloys compares favourably with previous results (see, for example, the values obtained by Meyer (1952) for Cu–Ni alloys at a wavelength of 3 cm). The value of $2{\cdot}22 \pm 0{\cdot}02$ recently reported by Bagguley and Harrick (1954) is clearly in agreement, within experimental error. In their experimental arrangement the specimens were very small spheres which were thus uniformly magnetized, while in the present experiments with plane specimens the finite skin depth caused a non-uniform magnetization. Kittel and Herring (1950) considered the effect of non-uniform magnetization on the position of the resonance line and predicted a shift of only a few oersteds. Their predictions have been criticized by Rado and Weertman (1954), but shifts of 50 oersteds could barely be detected with certainty with the present experimental accuracy.

4.2. *Line Widths*

The values of $1/T_2$ recorded in tables 1 and 2 refer to polycrystalline materials and therefore include the artificial broadening of the line known as 'anisotropy' broadening. With a single crystal, the resonance field $H_z^{\max}$ is dependent upon the angle between the steady field and the crystal axes. Thus, in a polycrystalline specimen with a random orientation of these axes relative to H_z, a range of resonance fields will occur and produce an apparent increase in line width. Preliminary measurements on a single crystal of nickel indicate that this factor increases $1/T_2$ by about 25% at room temperature, while at 200°c an increase cannot be detected with certainty. The anisotropy broadening is proportional to K_1/M where K_1 is the first order anisotropy constant, and from the data available (Bozorth 1951) it appears that the relative increase in $1/T_2$ in the Ni–Cu alloys should be similar to that in pure nickel. At the higher temperatures the broadening is expected to be small. Anisotropy constants for the other alloys studied are not known, but they are unlikely to differ drastically from those found in the Ni–Cu alloys.

For a given alloy, the shape of the curve of $1/T_2$ against temperature is similar to that found by Bloembergen (1950) for nickel, when comparison is made over a corresponding range of reduced temperature. The curve of $1/T_2$ against reduced temperature is not unique and varies from alloy to alloy.

It appears however that a single curve represents the results when $1/T_2$ is plotted against saturation intensity of magnetization, $1/T_2$ increasing as the intensity falls. Such a curve is shown in figure 4 in which the results for nickel, Ni–Cu, Ni–Al and Ni–Sb alloys are plotted. In the interest of clarity a few points have been omitted; these fall within the limits given by existing points. The mean line which has been drawn obeys an exponential law quite closely.

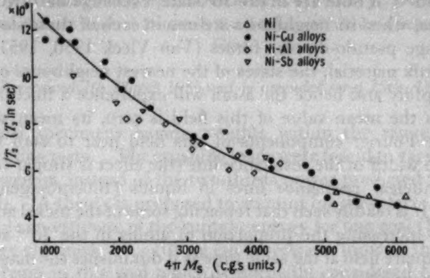


Figure 4. The variation of $1/T_2$ with saturation induction ($4\pi M_s$).

The results for the Ni–Mn alloys also fall reasonably close to the curve of figure 4, usually, in fact, within the limits of experimental error. Such data have been omitted from the figure, however, because Ni–Mn alloys, with two dissimilar magnetic carriers, differ from the other alloys studied. This is clearly demonstrated by the g values, and an examination of tables 1 and 2 suggests the the same might be true of the values of $1/T_2$.

Both experimental error and anisotropy broadening introduce some uncertainty, but the trend of the results is clear. Within experimental error it may be said that, in the present experiments, the line width specified by the damping constant $1/T_2$ is determined by the magnetization M and is independent of the constitution and temperature of the material.

It will be seen from the figure that $1/T_2$ changes only slowly with magnetization in the region of $4\pi M = 6000$ e.m.u. cm^{-3}. The value of $4\pi M$ for nickel at 0°K is only 6400 e.m.u. cm^{-3}, and one is thus tempted to deduce that the line width in nickel at this temperature should not differ markedly from that found at room temperature. No measurements of $1/T_2$ for metallic single crystals at low temperatures appear to have been reported. Yager (quoted by Kittel and Abrahams 1953) found that the line width in supermalloy was roughly constant between room and liquid air temperatures, while an almost constant line width in nickel between room and liquid hydrogen temperatures is believed to be consistent with the measurements of Bagguley and Harrick (1954).

A satisfactory theoretical explanation of these results cannot yet be given. Various mechanisms have been suggested and discussed (without reaching any agreed conclusion) by, for example, Bloembergen (1950), Van Vleck (1951), Kittel and Abrahams (1953) and Bloembergen and Wang (1953, 1954). The suggested mechanisms include spin–spin and spin–lattice relaxation, eddy current damping, and interaction with conduction electrons. In their discussions these authors have mainly considered the temperature dependence of the resulting line width; the present investigation indicates that in this case at least a dependence on magnetization should be sought and not primarily a dependence on temperature.

Stevens (unpublished) has suggested a possible mechanism which has the merit that the required dependence on magnetization could follow qualitatively fairly naturally. In nickel the observed magnetic moment might be considered to indicate the relative numbers of atoms with nine and ten electrons in the 3d shell.

Consider a linear chain of such atoms. A particular atom with magnetic moment one Bohr magneton will experience no resultant field if both of its neighbours are in the state $3d^{10}$: if both are in the $3d^9$ state, exchange narrowing should take effect. However, when its neighbours are one in each of these states, a field due to the short range pseudo-dipolar forces (Van Vleck 1950, 1951) will become apparent. In bulk material, the states of the nearest neighbours of a given atom will fluctuate rapidly and hence the atom will experience a fluctuating magnetic field. Although the mean value of this field is zero, its mean square value is positive and the Fourier components of this field near to zero frequency will contribute to the width of the resonance line (the effect is similar to the motional effects on the nuclear resonance lines in liquids (Bloembergen, Purcell and Pound 1948)). It is readily seen that replacing some of the nickel atoms by copper has the effect of increasing the proportion of atoms in the $3d^{10}$ state: this will affect the mean square field at the same time as it diminishes the magnetic moment. In a very qualitative manner, one can see that if the alloy contains few atoms in the $3d^9$ state (low nickel content) the mean square field will be small and the line width also. For pure nickel, exchange narrowing should produce a relatively narrow line and between these two limits the line width must rise and fall again. In the region experimentally investigated (100%–75% Ni) an increase in line width with decrease in magnetization could be expected. The model is much simplified—for example, atoms in the state $3d^8$ have been ignored—and a quantitative estimate of the line width to be expected may well prove difficult.

ACKNOWLEDGMENTS

The authors wish to record their thanks to Professor L. F. Bates for his interest and advice during the course of this work. They are also indebted to Dr. G. D. Adam and Dr. K. W. H. Stevens for many helpful discussions. In addition K. H. R. thanks the British Council for a scholarship which enabled him to participate in this research.

REFERENCES

ADAM, G. D., and STANDLEY, K. J., 1953, *Proc. Phys. Soc.* A, **66**, 823.

BAGGULEY, D. M. S., and HARRICK, N. J., 1954, *Proc. Phys. Soc.* A, **67**, 648.

BENSON, F. A., 1953, *Proc. Instn Elect. Engrs*, **100**, 85.

BLOEMBERGEN, N., 1950, *Phys. Rev.*, **78**, 572.

BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73**, 679.

BLOEMBERGEN, N., and WANG, S., 1953, *Tech. Rep. No.* 188 *Cruft Laboratory*; 1954, *Phys. Rev.*, **93**, 72.

BOZORTH, R. M., 1951, *Ferromagnetism* (New York: van Nostrand).

GRIFFITHS, J. H. E., and MACDONALD, J. R., 1951, *J. Sci. Instrum.*, **28**, 56.

HOSELITZ, K., 1952, *Ferromagnetic Properties of Metals and Alloys* (Cambridge: University Press).

KITTEL, C., 1947, *Phys. Rev.*, **71**, 270; 1948, *Ibid.*, **73**, 155.

KITTEL, C., and ABRAHAMS, E., 1953, *Rev. Mod. Phys.*, **25**, 233.

KITTEL, C., and HERRING, C., 1950, *Phys. Rev.*, **77**, 725.

MEYER, A. J. P., 1952, *C.R. Acad. Sci. Paris*, **235**, 1382.

MONTGOMERY, C. G., 1947, *Technique of Microwave Measurement* (New York: McGraw-Hill).

OSBORNE, J. A., 1945, *Phys. Rev.*, **67**, 351.

RADO, G. T., and WEERTMAN, J. R., 1954, *Phys. Rev.*, **94**, 1386.

VAN VLECK, J. H., 1950, *Phys. Rev.*, **78**, 266; 1951, *Physica*, **17**, 234.

YAGER, W. A., GALT, J. K., MERRITT, F. R., and WOOD, E. A., 1950, *Phys. Rev.*, **80**, 744.

YAGER, W. A., and MERRITT, F. R., 1949, *Phys. Rev.*, **75**, 318.

The Apparent Strength of Extensively Cracked Materials

By D. J. MILLARD, P. C. NEWMAN†, and J. W. PHILLIPS
National Coal Board, Central Research Establishment, Stoke Orchard, Cheltenham, Glos.

MS. received 7th March 1955 *and in amended form 25th April* 1955

Abstract. Coal specimens having weights within the range 10^{-9}–10^{3} g were broken in crushing tests. The crushing loads were found to vary as the half power of specimen weight, instead of the two-thirds power predicted by simple dimensional analysis. A theory is proposed to account for these facts and this theory is used to provide an estimate of the rupture energy for one coal (Llandebie). The value of energy obtained, $3{\cdot}0 \times 10^{4}$ erg cm^{-2}, seems to indicate that plastic deformation is occurring in this coal which is normally considered completely elastic. The occurrence of plastic flow in very small particles of any coal under similar conditions of loading is a logical sequence of the theory. This plastic flow has been observed in two of the coals used in this work.

§ 1. INTRODUCTION

It is well known that strengths of materials as measured in various standard tests are anomalous, in that the applied load per unit area at failure of geometrically similar specimens varies with the dimensions instead of remaining constant, as predicted by simple dimensional theory. The smaller specimens appear relatively stronger. To explain this it is necessary to suppose that the material possesses features which do not scale geometrically. Structural weaknesses and flaws of one kind or another have been invoked, and theories based on maximum sizes, or on frequency distribution of sizes, of flaws have been advanced. Those by Griffith (1924) and Freudenthal (1946) are typical. No theories have as yet been generally accepted, nor promise to be generally applicable. Results are given in this paper of experiments on the failure under compressive loads of coals containing many cracks and planes of weakness. A theory is advanced to account for the observed results.

§ 2. EXPERIMENTAL WORK

Measurements of crushing load have been made on three coals, of which details are given in table 1. The coals were chosen to cover a wide range of rank in British coals ; they were Llandebie and Langwith, which are extremes, and Betteshanger,

Table 1. Details of Coals Used (percentage composition)

Colliery and Seam	Llandebie Lower Pumpquart	Betteshanger H	Langwith High Main
Moisture	1·6	0·9	9·2
Ash	0·8	4·8	4·6
Volatile matter‡	4·5	18·7	39·9
Carbon matter‡	92·6	90·8	81·5
Hydrogen‡	2·8	4·6	5·2

‡ Dry ash free.

---

† Now at the Mullard Research Laboratories, Salfords, Redhill, Surrey.

which is representative of the intermediate coals of low porosity. All three coals provided lumps which were extensively cracked. The lumps of Betteshanger and Langwith were of irregular shape. From the fragments of a broken specimen it was possible to select a lump large enough to serve as the next specimen. Generally, there was only one such lump, the rest of the previous specimen having split up into small pieces. The Llandebie was cut into cubes, the preparation of which was hampered by the cracked state of the coal. Many of the specimens fell apart before completion and had to be discarded. For this coal, specimens of each different size were cut from the same parent lump.

Two sets of equipment were used for the crushing tests, one dealing with lumps weighing 10^{-2}–10^{3} g and the other with particles weighing 10^{-9}–10^{-5} g. The larger lumps were crushed between the platens of a press ; three different systems of loading the cubes were used, one platen being either a plane, cylinder or sphere, while the second platen was always a plane. The radii of the cylinder and sphere were scaled linearly with the edge of the cube of coal. The smaller particles were crushed between glass platens under a microscope, in order to facilitate observation of the particles during crushing. The weights of these particles were computed from the observed dimensions, by cubing the geometric mean of maximum and minimum diameters and assuming a density of 1·3 g cm^{-3}. Visual observation was essential, as Boddy (1943) has shown that very small particles of coal may flow instead of shattering. Some of the smallest particles in these experiments did flow, and these results have been omitted from the analysis described below, but they are mentioned later. Details of the apparatus for crushing the small particles will be published elsewhere.

The results of these experiments are summarized in figures 1–3, in which crushing loads are plotted against specimen weights on logarithmic scales for each coal. The number of readings which each point represents is shown on the figures ;

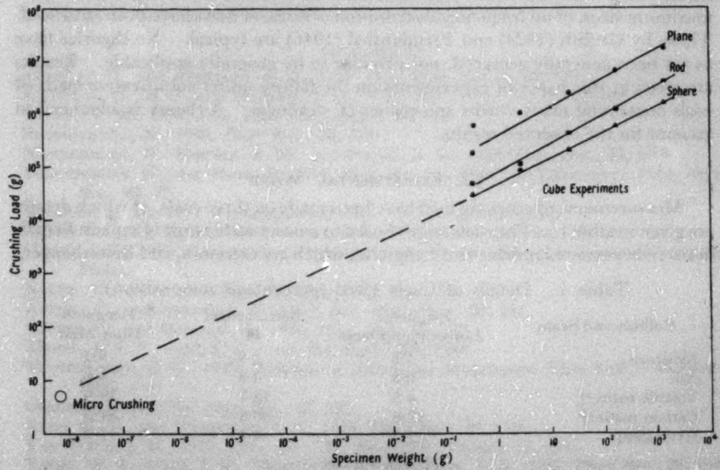


Figure 1. Crushing load as a function of specimen weight for Llandebie anthracite.

the lines are obtained by regression analysis on the total of results. In the case of Llandebie cubes (Burton and Phillips 1952) it is not clear which of the three lines

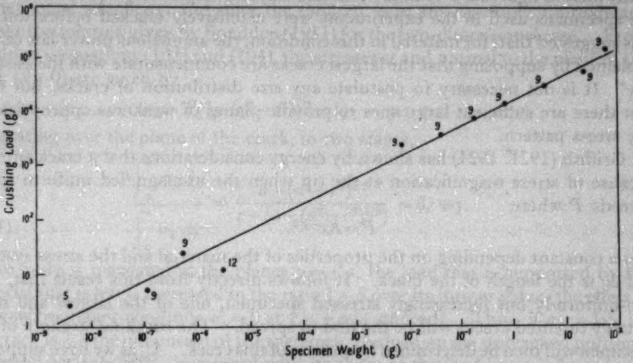


Figure 2. Crushing load as a function of specimen weight for Langwith coal. The number of observations at each point is shown.

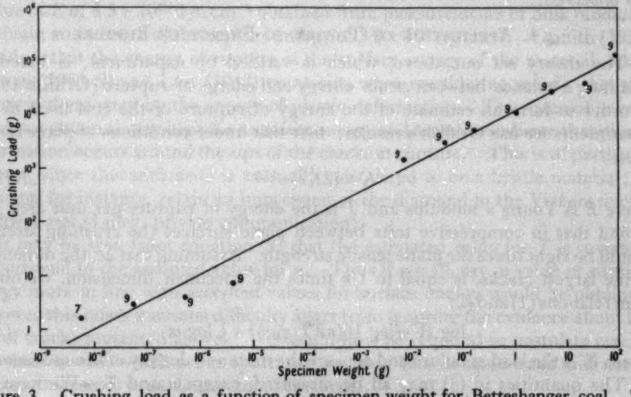


Figure 3. Crushing load as a function of specimen weight for Betteshanger coal. The number of observations at each point is shown.

should be extrapolated to include the results on irregular particles of micron sizes. The power laws relating crushing load to weight for each coal are shown in table 2.

Table 2. Power Law relating Crushing Load to Specimen Weight

Coal	Llandebie	Betteshanger	Langwith
Power	0·52	0·51	0·49
Error	±0·02	±0·05	±0·04

§ 3. PROPOSED THEORY OF APPARENT STRENGTH

For a homogenous material dimensional analysis requires the power laws of the lines in figures 1–3 to be 2/3; in fact they are very closely equal to 1/2. All the specimens used in the experiments were extensively cracked before testing. It is suggested that, for material in this condition, the anomalous power law can be explained by supposing that the largest cracks are commensurate with the particle size. It is not necessary to postulate any size distribution of cracks, but only that there are sufficient large ones to provide planes of weakness appropriate to any stress pattern.

Griffith (1921, 1924) has shown by energy considerations that a crack spreads because of stress magnification at the tip when the unmagnified uniform stress exceeds P where

$$P = Kc^{-1/2}. \qquad \ldots\ldots(1)$$

K is a constant depending on the properties of the material and the stress system, and $2c$ is the length of the crack. It follows directly from this result that, in a non-uniformly but increasingly stressed specimen, one of the largest and most suitably oriented cracks will be the first to spread. The breaking strength of the specimen will then be determined by the size of this crack. If, as we have supposed the largest cracks are commensurate with the size of specimen, then the applied load per unit area for breaking must vary inversely as the square root of linear dimension. Thus, the applied load will vary as the square root of specimen weight, which is in fact observed.

§ 4. APPLICATION OF THEORY TO ENERGY OF RUPTURE

The theory set out above, which is verified by experiment, is based on obtaining a balance between strain energy and energy of rupture (Griffith 1921). In order to form an estimate of the energy of rupture of the coal used in our experiments we use Griffith's results; first that under conditions of plane tensile stress

$$K = (2ET/\pi)^{1/2} \qquad \ldots\ldots(2)$$

where E is Young's modulus and T is the energy of rupture per unit area, and second that in compressive tests between plane surfaces the crushing strength should be eight times the plane tensile strength. Assuming that $2c$, the dimension of the largest cracks, is equal to $1/n$ times the specimen dimension, we obtain from relations (1) and (2)

$$\log W = \log 16(nET/\pi\rho)^{1/2} + \tfrac{1}{2}\log m \qquad \ldots\ldots(3)$$

where W is the load at failure and m, ρ are the mass and density of the specimen.

The quantities in (3) may all be measured, except n and T. We have put $n = 1{\cdot}5$ in the calculations below, which implies that the cracks are nearly co-extensive with the lump. The justification for this is principally the difficulty, which has been outlined above, in obtaining specimens. When attention is turned to coals from which it is easy to obtain cubic, or other arbitrarily shaped, specimens and which presumably do not contain extensive cracks, then the breaking load and specimen weight are no longer connected by a half-power law (Walton, private communication).

Griffith's theory applies to a crack in an infinite body and its application to a crack which is nearly coextensive with the body requires justification. The solution for a finite body is not available, but an indication of the error in truncating an infinite solution can be obtained; this may be done by considering what proportion of the extra load on the surrounding material, due to the presence of the crack, is borne by the material in the immediate vicinity. In the interest of mathematical simplicity, and since the estimate is in any case only approximate, we take the solution given by Sneddon (1951) for the two-dimensional case. From Sneddon's equations (102) and (104), the stresses at and normal to the plane of the crack ($x=0$) are given by

$$\sigma_x = P_0\{y/(y^2-c^2)^{1/2}-1\} \qquad y>c.$$

Integrating over the plane of the crack, in two stages,

$$\frac{\int_a^\infty \sigma_x\,dy}{\int_c^a \sigma_x\,dy} = \frac{a-(a^2-c^2)^{1/2}}{c-\{a-(a^2-c^2)^{1/2}\}} = \beta, \text{ say.}$$

If the body is truncated at the planes $y = \pm a$, the load that is supported by the material that is left is increased in the ratio $1+\beta$ over its former value (the effect of the new boundary condition $\sigma_y = 0$ at $y = \pm a$ is ignored). Thus, if $a/c = 1{\cdot}5$, i.e. the crack is 2/3 the dimension of the specimen, the load on the uncracked portion is increased by only 60%, compared with its share of the load in an infinite body. This is regarded as a sufficient justification for applying Griffith's theory to an extensively cracked material.

Using the results from cubes of Llandebie compressed between two planes, and a value of E of $8{\cdot}5 \times 10^{10}$ dyn cm^{-2} obtained from measurements of bulk modulus, we obtain an estimate of the rupture energy of $3{\cdot}0 \times 10^{4}$ erg cm^{-2}. Griffith (1921) considers that the energy of rupture is the surface energy of the material, while Orowan (1948–9) and Low (1943) point out, when considering metals, that this energy is dominated by the energy of plastic deformation. The value of energy which we obtain is very high for a surface energy and seems to indicate that plastic deformation occurs around the tips of the cracks at rupture. This is of particular interest, since this anthracite is normally considered to be a brittle material; it does not, for instance, retain an impression of the diamond in the Vickers microhardness test (Lilley, private communication).

It may be seen from equation (3) that the estimated value for T is inversely proportional to the assumed value for n. Thus if $n = 10$, say, a value of rupture energy more in line with accepted values for surface energy would be obtained. However this raises a serious difficulty apart from ignoring the evidence about the size of cracks presented above. We would now be required to postulate cracks, for which $n \sim 10$, which are proportional to lump size in some coals and not in others; this seems to us the greater of two evils.

§ 5. Applications of Theory to Very Small Sizes

The theory proposed above predicts that the apparent crushing strength of particles will continue to increase with decreasing particle size, the upper limit being set by the tensile strength of the uncracked material. Energy considerations, however, suggest that in uncracked material, the shear yield stress will be less than the tensile rupture stress and will be reached first as the particle size decreases. Thereafter, below a certain particle size, the cracked material will flow instead of shattering. This is the phenomenon observed by Boddy (1943), by us in the present experiments, and by others (Bangham and Berkowitz 1947).

Acknowledgment

The authors express their thanks to the National Coal Board for permission to publish this paper; the views expressed are those of the authors and not necessarily those of the Board.

References

Bangham, D. M., and Berkowitz, N., 1947, *Research, Lond.*, **1,** 86.

Boddy, R. G. H. B., 1943, *Fuel, Lond.*, **22,** 56.

Burton, L. D., and Phillips, J. W., 1952, N.C.B. Scientific Dept. C.R.E.I. Report 154.

Freudenthal, A. M., 1946, *Proc. Roy. Soc.* A, **187,** 416.

Griffith, A. A., 1921, *Phil. Trans. Roy. Soc.* A, **221,** 163; 1924, *Proc. 1st Int. Congr. for Appl. Mech.*, p. 55.

Low, S. R., 1953, *Relations of Properties to Microstructure* (Cleveland: Amer. Soc. Metals), p. 163.

Orowan, E., 1948–49, *Rep. Progr. Phys.*, **22,** 186 (London: Physical Society).

Sneddon, I. N., 1951, *Fourier Transforms* (New York: McGraw-Hill).

On the Automatic Design of Optical Systems

By G. BLACK

British Scientific Instrument Research Association, Southill, Chislehurst, Kent †

MS. received 11*th February* 1955 *and in amended form* 19*th May* 1955

Abstract. Electronic digital computers in Britain and the U.S.A. have been used for the tracing of rays at high speeds through optical systems. The larger machines have enough auxiliary storage, either magnetic drum or tape, to accommodate enormous amounts of information so that it is possible to carry out extensive numerical investigations concerning lens systems.

This paper deals with machine methods of minimizing the calculated geometrical aberrations of systems in much the same way as the optical designer using conventional desk machines. It will be shown that the main designer's task can be achieved automatically in a remarkably short time. The machine referred to is the Manchester Electronic Computer (MADAM), but the remarks apply to any high speed machine with a reasonable storage capacity. It is not proposed to enter into any of the programming details, these would require a separate paper.

§ 1. Introduction

The general method of design of an optical system usually begins with a gaussian investigation which determines from a knowledge of the speed required and field sizes a rough arrangement of powers of lenses. A first order or Seidel investigation determines convenient shapes and distributions of lenses and lens powers, so that the Seidel residuals are reduced to zero or to suitably small values.

For some purposes, where the necessary angular fields and apertures are small, this arrangement may be good enough. For more ambitious systems, i.e. systems of large aperture or field size, an extensive adjustment of variable parameters is required whereby the higher order aberration residuals are balanced so that good performance is obtained over the required aperture and field. With desk machines this latter process is a matter of experience, although recently differential methods of fine correction have reduced the need for precise ray tracing somewhat.

Essentially, the design problem (after gaussian requirements are satisfied) is that of the minimization of a series of aberration functions which depend on n variables, n being the number of variable constructional parameters. The problem is a special one however in that the variable-function system is largely ill-conditioned and non-linear. For an average photographic lens system, e.g. the 'Cooke' anastigmat or the 'Tessar', it is possible with the aid of a fast electronic computer to determine a set of aberration residuals in less than 1 minute. Such a set of residuals might consist of some 16 numerical values composed as follows: chromatic aberration (3 values), axial spherical aberration (3 values), coma, distortion, oblique spherical aberration, and astigmatism for (say) two field angles (two sets of 5 values each).

---

† Extra-mural research carried out at the Computing Machine Laboratory, Manchester University.

In the example given later, axial chromatic aberration is defined simply as the distance between the intersection points of C and F rays and the optical axis. Axial spherical aberration is calculated by tracing finite meridional rays parallel to the optical axis at different apertures, and is given as the distance between the intersection points of the emergent rays with the optical axis and the gaussian focal plane position. The tangential coma is defined as η_Q/η_P-1 where η_P and η_Q are the respective heights of the emergent principal ray and the intersection point of the upper and lower rays measured along the perpendicular dropped from this point to the optical axis. Distortion, usually measured as a principal ray aberration in the gaussian image plane, has been calculated as $\eta_P/\tan U_p-1$ where U_p is the angle between the emergent principal ray and the optical axis, the equivalent focal length of the system being 1 inch. The sagittal and tangential focal distances given are those between the feet of perpendiculars dropped from the respective foci to the optical axis and the gaussian focal plane position. Oblique spherical aberration is defined as the axially projected distance between the intersection point of the upper and lower rays and the tangential focus.

A set of aberration residuals such as the one described above defines (to the optical designer at least) the usefulness of the system for his purpose. The problem now is to reduce these residuals employing any variable or combination of variables of the system within certain limits.

§ 2. PERFORMANCE NUMBERS

On seeing a set of aberration residuals the designer automatically picks out the most undesirable, and proceeds to reduce this until some other residual assumes importance, at which point the minimization operations are changed to deal with the new situation. Using electronic machines it is convenient (though not essential) to have a single function ϕ by which the system performance is assessed; such a number can be obtained simply by forming a sum of the squares of the residuals. Another function can be obtained by finding the average modulus of a set of residuals. A function which has been used with some success at Manchester is that formed by summing the squares of the maximum moduli of given groups of residuals with appropriate scaling or weighting factors.

It is obvious that performance numbers of this sort have no simple optical significance, they relate simply to the state of the residuals at a point in the minimization process. The precise nature of the function ϕ does of course determine the distribution of residuals at the conclusion of any minimization process but a vital feature of the function ϕ is that its nature becomes less important as the residuals become smaller. Were it possible to reduce ϕ to zero, then all the residuals would be zero. It is true that we are concerned here with geometrical aberration functions and we ignore the fact that the optical performance of a system which has an extremely low value of ϕ may be slightly worse in fact than another which has a slightly higher value. Using such a function the optical design problem is reduced to an extensive numerical analysis of an n-dimensional surface, a problem which is by no means confined to optics. A considerable amount is already known about reducing the aberrations of a lens system; for instance the designer knows well how powerful is the method of lens bending, and how necessary it is to be able to control a particular residual with respect to the other residuals. These operations have almost identical parallels in the relaxation technique described in the work which follows.

The method of 'steepest descent' is one which can be applied to the problem of reducing a function of n variables, but the author has found that this method has limitations when the value of n is large, which is generally the case in optical design. It appears that the reward of this method, i.e. the rapidity of reduction of the performance number, is not sufficient compared with the machine time required for the calculation of the most desirable path: it may well be that for a few variables (5 or 6) the method would give useful results. A serious disadvantage of this latter method is that the calculation of the most desirable path must be done first, before any changes in the system can be made. Using 15 or 20 variables this time may be half an hour, and if a machine fault occurs in this period, most, if not all the work may be lost. In the relaxation technique now to be described an improvement in the performance number can occur in less than 2 minutes, and once it has been obtained the resulting system can usually be extracted from the machine unless the fault is so disastrous as to erase the relevant information regarding the improved system from the surface of the magnetic drum.

§ 3. MINIMIZATION PROCESSES

3.1. *Variable by Variable Minimization*

The simplest way of reducing ϕ is to alter each variable of the system in turn, recalculating the aberration residuals for each new value of the parameter in question. If, say, the surface radii are taken as a block of variables, the machine traces a set of rays through the system storing all the inter-surface information so that the traces can be 'picked up' at any point within the system. One of the surfaces can now be altered (by adding or subtracting a small increment to the surface curvature) and the new value of ϕ obtained very rapidly. If this value is greater than the original value, the sign of the increment is changed and ϕ recalculated. Each time a smaller value of ϕ is obtained the new parameter value is retained and control moves on to the next variable in the block. If no fall in ϕ is obtained in either case then the function is at, or very near to, a minimum for that parameter, so the original value of the parameter is retained and control moves on as before. This is the basis of the variable by variable process which can be carried out at high speed; the process is extremely valuable as far as it goes, and can be regarded as the basic unit operation in automatic design. On a seven-surface 'Tessar' lens this operation can be carried out using each of the curvatures in about 15 minutes. The speed of reduction of ϕ depends on the initial distribution of the residuals, and on the magnitude of the increments, but it is a simple matter for the machine to find a suitable increment should this be necessary. A simple means of doing this is to count, in a particular storage line, the number of times a different constructional parameter is retained, i.e. without improvement in the ϕ value. When this number exceeds a given proportion of the total number of variable parameters (say 60%), the increment value can be automatically halved. The reverse process is equally important when the increment is too small. This ' step length ' or increment variation technique is widely used in automatic numerical integration.

The variable by variable method is iterative, and having run through a set of variables once, control can revert to the first variable in the series, the refraction conditions on the second run usually being different from those on the first run. In some systems a point can be reached where no further reduction in ϕ is possible unless the value of the parameter increment is reduced, and in any case it will always be necessary to apply other minimization processes to ensure that no smaller value of ϕ can be obtained. The variable by variable method generally leads to an optical system which has a smaller value of ϕ than the original system but with the state of affairs that no improvement can be achieved on any single variable. Generally, minimization by this method does not alter the equivalent focal length of the system very appreciably; this is largely because some accepted parameter changes shorten the focal length while others lengthen it. After the variable by variable process has been employed using a particular set of variables it is very necessary to step the system up or down to the original focal length, and to ensure that the rays, the configuration of which determine ϕ, are passing through the system in the correct manner. This means that the nodal points must be calculated so that the off-axis rays are defined in a constant manner. In all the work which follows this precaution is taken after every change in the system.

3.2. *Block Operations*

What is required at the close of the variable by variable process is a method of altering the refraction conditions associated with each variable while the general composition of the system is appreciably unaltered, so that the variable by variable method may continue with advantage. Perhaps the most popular relaxation device used by the designer is that of ' lens bending ', the purpose of which is that of altering the distribution of aberration residuals while the overall focal length and colour corrections remain fixed. Lens bending can be likened to a block operation of relaxation, it being affected by adding (or subtracting) equal increments to the curvatures of a given lens, the overall power remaining constant. In optics however the increment is chosen with the particular purpose of improving the overall distribution of residuals. Automatic operations of this sort can be carried out on from two to n variables with two purposes in mind: (i) to reduce the latest value of ϕ directly, (ii) to change the refraction conditions associated with each variable. At the completion of one or a series of block operations control can continue with the variable by variable method or proceed to another minimization process. A particularly useful block operation which has been used with success at Manchester is that of adding or subtracting small equal increments to all parameters. The effect of this operation varies considerably from system to system, but for photographic lenses with the diaphragm position somewhere near the middle of the system, changes of this sort result in violent variations in the residuals while the equivalent focal length is not greatly affected. Block operations are iterative within themselves, and the machine can be programmed to determine a suitable increment or series of increments to suit the variations produced in the performance function.

3.3. *Group Operations*

In some cases the variable by variable method followed by the block relaxation process leads to a system which has a set of residuals all less than those of the original or some intermediate system. It is a simple matter in these circumstances to carry out a group operation, which is in effect an extrapolation from two systems, to reduce ϕ. Group operations can always be carried out from two systems to a third system having a smaller value of ϕ unless the present best system is at a true topographical minimum. Group extrapolations are extremely valuable for the following reason. Since the number of variables changed in group operations is considerable and the individual increments generally small, the relation between the extent of the operation and the change in the residuals is largely linear. This fact allows one to 'forecast' a set of residuals without actually calculating them by precise ray tracing, provided that all 'reference' systems have constant focal length, and an exactly similar ray configuration. It is possible to carry out several group operations in two or three seconds with appreciable decreases in ϕ after each one. Group changes require that as many system specifications together with their calculated residuals be held in the machine as possible. It is often the case that some intermediate system which has been discarded because of its high ϕ value is eminently suitable for a group operation with the latest, much better system. The process is iterative in that increasing values of the particular increments may be applied until the value of ϕ begins to increase once more. Control can now revert to the block relaxation process or to the variable by variable method already described.

3.4. *Random Operations*

Up to now we have discussed only changes in single variables, or all variables, or groups of variables, i.e. surface radii, or axial displacements. When ϕ will not reduce by any of the foregoing methods it is sometimes possible to obtain a further reduction by random selection of variables. The Manchester machine has a random number generator by which a series of twenty random binary digits are produced by a noise generator. There are several instructions named 'logical instructions' which are used in this connection, being instructions which have no relation to the interpretation of rows of digits as numbers. These instructions can be used to select say the middle five binary digits from the row of twenty, enabling one to obtain a series of random numbers from zero to 31. Using this facility it is possible to 'sprinkle' the present best system (of say 2^5 variable parameters) with a random selection of increments. The randomness can apply not only to the variable parameters but also to the size of the increment. If the present best value of ϕ is already low the tendency of this random sprinkling process is to increase ϕ. If however ϕ is reduced the new system is accepted and control continues the random 'search' or moves to some other minimization process.

The main advantage of increasing ϕ by random methods is that systems can be evolved for use in the group relaxation process already mentioned; successful group operations requiring two systems, one having quite small residuals and another (randomly produced in this case) having residuals larger, but of the same signs as those of the first system. Random operations can thus be used to verify that the present best system is probably the best which can be obtained in a particular case.

§ 4. THE METHOD WITH AN EXAMPLE

The flow diagram of an automatic design programme which has been used successfully at Manchester is shown in figure 1. The arrows show how successive iterations can be conducted. Though simple enough in structure, a programme of this sort is vastly complicated by necessary organization within the machine. A major part of the programming concerns the locations of ray-tracing data, systems, constants and blocks of residuals. The Manchester machine has an available magnetic drum storage of some half million binary digits; this amount can be regarded as quite adequate to deal with systems as complex as those usually met with by the designer. The problem of physical limitation is complicated by the fact that one minimization process may make, say, an axial separation negative, while the next may make it positive and sensible again, the value of ϕ falling throughout these changes. The operator must interfere occasionally, although this means only a pause of seconds, while the storage line, in which is contained the number of the variable concerned, is altered by manual instruction.

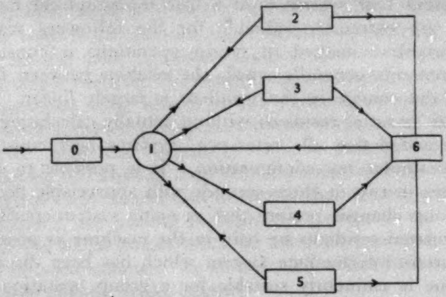


Figure 1. Flow diagram of design programme. 0, read in system constants, aperture ratio and field sizes; 1, control (the master routine); 2, variable by variable process; 3, block operations; 4, group operations; 5, random operations; 6, output.

It is proposed now to give a simple example of machine minimization. The following first set of residuals are those of a well-known 'Tessar' lens which has been in use for many years. The residuals on the left are the original values: those on the right are the residuals of a system calculated after some two hours variable by variable minimization followed by four block and four group operations, requiring 2½ hours in all. The transverse chromatic aberration has not been included in the performance function chiefly because no changes either in refractive indices or dispersions were contemplated. In general, this important aberration should of course be calculated in the usual way and its

	Focal length 1 in.		
	C−F	−0·0002	+0·0003
	Axial spherical aberration	−0·00038	+0·0072
		−0·00451	+0·00017
		−0·00486	−0·0017
	Coma	−0·00571	+0·00330
	Distortion	+0·001	−0·00059
20°	Oblique spherical aberration	+0·0142	−0·00218
	t focus	+0·00046	−0·00547
	s focus	+0·0048	+0·00976
	Coma	−0·0200	−0·00203
	Distortion	+0·0045	−0·00091
30°	Oblique spherical aberration	+0·0856	+0·01606
	t focus	+0·0465	+0·01183
	s focus	−0·0184	−0·00924

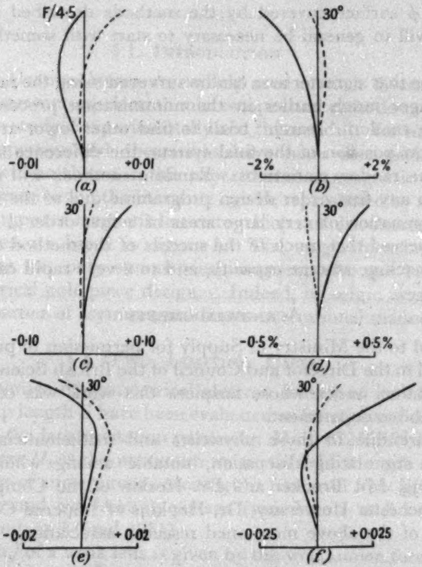


Figure 2. Graphs showing improvement in aberration curves.

Ordinates : (*a*) aperture, (*b*)–(*f*) field angle.

Abscissae : (*a*) longitudinal spherical aberration; (*b*) percentage tangential coma, (*c*) oblique spherical aberration; (*d*) percentage distortion; (*e*) sagittal curvature; (*f*) tangential curvature.

In diagrams (*a*) and (*c*) a negative sign signifies undercorrection, i.e. the foci concerned are short of the reference point or plane. In diagram (*d*) negative distortion is ' barrel ' distortion. In diagram (*b*) negative coma is ' inward ' coma. In diagrams (*e*) and (*f*) a negative sign signifies overcorrection, i.e. the foci concerned are beyond the reference plane.

value should form part of the performance number. The improvement achieved is shown diagrammatically in figure 2; the improved aberration curves are shown dotted, those of the original system in full. The values of the aberration residuals of the improved system were subsequently confirmed by ray tracing using an electrical desk machine. It is of interest to note that even this confirmatory computation requires more than 15 hours' work, as against the 2½ hours total time taken by the electronic machine to produce, as well as to confirm the improvement in performance.

§ 5. Conclusions

Automatic computing machinery which removes the great labour of ray tracing completely can be used to explore the possibility of improving existing lens types. The method described appears adequate for this purpose. It is clear that the final system produced by this technique depends to a certain extent on the distribution of residuals in the initial system. In other words the area of the ϕ surface covered by the methods described is by no means infinite, and it will in general be necessary to start with something like a first-order solution.

It may well be that a greater area can be surveyed using the random relaxation technique at stages much earlier in the minimization process. This would amount to using random ' search ' trials to find other lower areas, rather than to confirm the low position of the final system, the difference being in the size and extent of the random variations. Random methods will almost certainly be very useful in any first order design programme due to the necessity in this case for the examination of very large areas of a first order ϕ surface.

It will be observed that much of the success of the method described is due to a very large machine storage capacity, and to a very rapid calculation speed.

Acknowledgments

I am indebted to the Ministry of Supply for permission to publish this work and I am grateful to the Director and Council of the British Scientific Instrument Research Association under whose auspices this work was conducted under contract to the above department.

My thanks are due to those physicists and mathematicians with whom I have enjoyed stimulating discussion, notable among whom are the late Dr. A. M. Turing, Mr. Brooker and Dr. Hoskin of the Computing Machine Laboratory, Manchester University, Dr. Hopkins of Imperial College, London, and Mr. Bracey of the above mentioned research association.

A Unified Representation of Magnetic Electron Lens Properties†

By G. LIEBMANN
Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berks.

MS. received 1st *March* 1955 *and in amended form* 4th *May* 1955

Abstract. It is shown that the focal lengths of all known iron shrouded magnetic electron lenses lie between two fairly close 'limit-curves', if f is measured in the unit $S+D$ for symmetrical or $S+R_1+R_2$ for asymmetrical lenses of bore radii R_1 and R_2 and gap length S, f being plotted as function of $(NI)^2/V_r$. The lenses of the more usual S/D ratios ($0{\cdot}5 \leqslant S/D \leqslant 2$) lie very close to the lower 'limit-curve', i.e. all these lenses can be represented by a single focal length curve. A similar representation can be given for the other lens constants.

---

§ 1. INTRODUCTION

Magnetic electron lenses are widely used in such diverse apparatus as β-ray spectrometers, electron microscopes and cathode-ray tubes. They have been studied experimentally and theoretically for many years, and their focusing properties and aberrations have been discussed in numerous publications. It will be shown in this paper that most magnetic electron lenses which have been described in the literature can be represented by a *single focusing curve*. The main types of magnetic lenses to which this discussion will apply and their principal uses are listed in figure 1. The single focusing curve to be described embraces not only symmetrical magnetic electron lenses, but also lenses of asymmetrical pole piece design. Indeed, its origin was the desire to represent the properties of asymmetrical lenses in a rational manner.

§ 2. TWO LIMITING CASES

The focal properties of symmetrical electron lenses, of lens bore diameter $D=2R$ and pole gap length S have been evaluated as functions of the geometrical factor S/D, and of the lens excitation parameter $k^2=0{\cdot}022H_0^2R^2/V_r$, or alternatively $(NI)^2/V_r$ where H_0 is the maximum magnetic field strength along the lens axis, NI the number of ampere turns exciting the lens, and $V_r=V(1+10^{-6}V)$ the relativistically corrected accelerating voltage (van Ments and Le Poole 1947, Lenz 1950, Liebmann and Grad 1951).

The focal length f of a weak lens is given by the well known formula

$$f=45{\cdot}5V_r\Big/\int_{-\infty}^{\infty}H^2(z)\,dz.$$

In the extreme case $S/D\gg 1$, the graphical representation of $H(z)$ as a function of z is a nearly rectangular curve, of mean width S, and nearly uniform height

$$H(z)\simeq H_p=4\pi NI/10S.$$

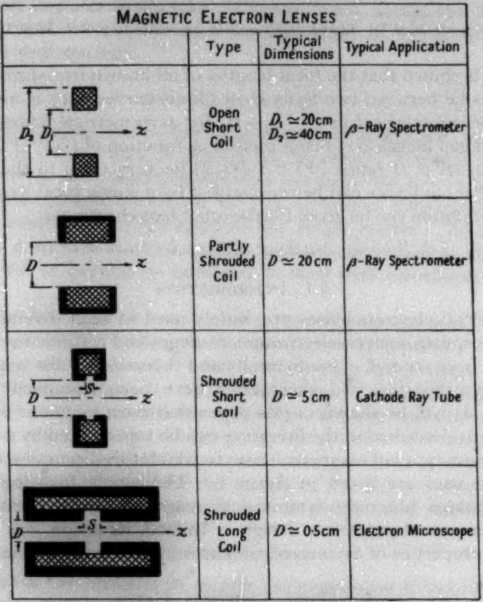
Hence in this extreme case the weak lens formula gives $f\simeq 28{\cdot}8\,V_rS/(NI)^2$.

---

† The results of this paper have been presented in part at the Physical Society Symposium on Electron Optics, in London in May 1953 and at the German Electron Microscopy Conference in Innsbruck in September 1953.

In the other extreme case, $S/D \ll 1$, $H(z)$ can be approximated by Ramberg's (1942) expression $H(z) \simeq H_0 \operatorname{sech}^2(2\cdot63z/D)$. Integrating,

$$\int_{-\infty}^{\infty} H^2(z)\,dz \simeq 0\cdot507\,H_0^2 D.$$



Magnetic Electron Lenses			
	Type	Typical Dimensions	Typical Application
	Open Short Coil	$D_1 \simeq 20$ cm $D_2 \simeq 40$ cm	β-Ray Spectrometer
	Partly Shrouded Coil	$D \simeq 20$ cm	β-Ray Spectrometer
	Shrouded Short Coil	$D \simeq 5$ cm	Cathode Ray Tube
	Shrouded Long Coil	$D \simeq 0{\cdot}5$ cm	Electron Microscope

Figure 1. Magnetic electron lenses.

The value of H_0/H_p becomes a linear function of S/D for small values of S/D (Liebmann and Grad 1951, fig. 7), i.e. $H_0 \simeq (4/3)H_p(S/D) = (16\pi/30)NI/D$. Hence in the second extreme case the weak lens formula leads to $f \simeq 31\cdot9V_r D/(NI)^2$.

§ 3. THE UNIFIED FOCAL LENGTH CURVE

The consideration of the two extreme cases $S/D \gg 1$ and $S/D \ll 1$ under 'weak lens' conditions showed that f is proportional to the lens diameter D in the one case and to the pole piece gap S in the other case. This suggests that f may be proportional to a linear combination of the two geometrical lens parameters D and S, and as the two numerical factors evaluated for the approximate formulae differ only by a small amount, this linear combination may be of the simple form $S+D$. The results of van Ments and Le Poole (1947), Lenz (1950), and Liebmann and Grad (1951), as well as additional unpublished data by the last-named authors, were therefore tested as to whether they could be expressed by a formula

$$f/(S+D) = \alpha V_r/(NI)^2 \qquad \ldots\ldots(1)$$

where α is a constant.

It was found that such a relation is indeed approximately satisfied, although the factor α is not a true constant. The range of values which the factor α can take is illustrated by figure 2, in which α is plotted as function of S/D for the two values $(NI)^2/V_r = 10$ and $(NI)^2/V_r = 100$; the smaller value of $(NI)^2/V_r$ corresponds to the weak lens condition considered earlier, whereas the larger one represents a fairly strong lens. The factor α increases as the lens excitation is increased, at first slowly, then more rapidly. It is also seen from figure 2 that the constant α is always smaller for lenses of more usual relative dimensions than for the two previously considered extreme cases, the minimum value, at $S/D \simeq 1$, being approximately 75% of the limiting value.

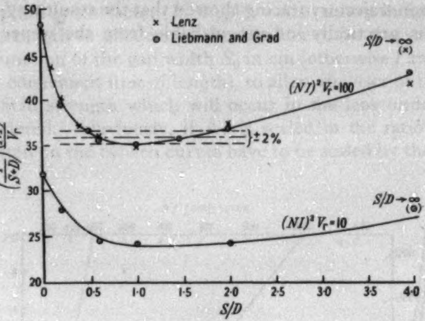


Figure 2.

It is therefore possible to enclose the data for *all* iron-shrouded magnetic electron lenses between two 'limit-curves' if $f/(S+D)$ is plotted as function of $(NI)^2/V_r$, as shown in figure 3. The full line curve corresponds to the smallest value of α, for each value of $(NI)^2/V_r$, whereas the broken curve represents the 'extreme' cases $S/D \gg 1$ and $S/D \ll 1$. The important fact is that the focal properties of the more usual electron lenses, for which $0{\cdot}5 \leqslant S/D \leqslant 2$, lie very close to the lower 'limit-curve', within about 2% to 3%. If a fairly accurate estimate of the focal length is desired for lenses which do not conform to this more common range of S/D values, the dependence of α on S/D shown in figure 2 can be used to correct the value of f taken from the lower 'limit-curve' in figure 3.

The lower 'limit-curve' in figure 3 splits at $(NI)^2/V_r \simeq 110$ into two branches. The lower branch represents the 'objective lens' condition (see Liebmann and Grad, fig. 11), in which the object being imaged by the lens is allowed to lie within the lens field (often met in electron microscope objectives). Along this branch, $f/(S+D)$ falls monotonically to a limiting value of $f/(S+D) \simeq 0{\cdot}088$ for $(NI)^2/V_r \to \infty$. The upper branch, shown by a broken line in figure 3, represents the 'projector lens' condition, which applies if the whole lens field is used in the image formation. In the latter case the lens reaches a minimum focal length, which is obtained at $(NI)^2/V_r \simeq 180$.

Figure 3 allows a further extension. It is known that the focal length of a magnetic lens consisting of a single circular wire loop of radius R is, under weak lens conditions, $f \simeq 98 V_r R/I^2$. It is also known (Lyle 1902) that the axial magnetic field of an ironless coil of inner radius R_1 and outer radius R_2 and square cross section of the winding of N turns can be represented by the magnetic field of a single circular wire loop of radius $R=\frac{1}{2}(R_1+R_2)$ carrying the current NI. Hence for such a coil $f\simeq 98\ V_r R/(NI)^2$. The relative axial field distribution of the open coil of square or nearly square cross section can be fairly well matched, at least over the more important part of the field where its strength is relatively high, by that of an iron-shrouded coil of ratio $S/D\simeq 0{\cdot}5$. The focal length of the square-section open coil can therefore be expressed by the relation (1), where the factor α has under weak lens conditions the value 32·7, which is quite close to the value for the extreme case $S/D\to 0$ of the iron-clad lens. For higher lens currents, where the weak lens condition is no longer satisfied, a numerical computation of the focal length by electron trajectory tracing showed that the resulting $f/(S+D)$ curve for the open coil is practically indistinguishable from the upper 'limit-curve' in figure 3.

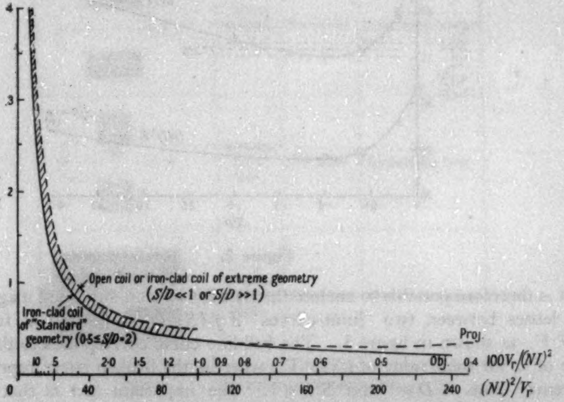


Figure 3. Unified focal length curves of magnetic electron lenses.

The following comparison with measured data shows how close the agreement with the curves of figure 3 can be. Becker and Wallraff (1938) investigated an ironless coil of inner diameter 9·0 cm, outer diameter 15·1 cm and length 5·4 cm, with 2750 turns. At an accelerating voltage of $V=10^4$ volts and a coil current of 0·220 A, a focal length of 16·0 cm was measured. The value taken from figure 3 for $(NI)^2/V_r=36{\cdot}5$ is $f/(S+D)=0{\cdot}90$, hence $f_{\rm calc}=16{\cdot}3$ cm. For a coil current of 0·325 A and an accelerating voltage of 15 kv, i.e. $(NI)^2/V_r=53{\cdot}3$, the measured f value is 12·5 cm and the calculated value 12·3 cm.

Similarly good agreement with figure 3 is shown by the data given by Siegbahn (1944) for his β-ray spectrograph lens, which is partly shrouded. From Siebahn's figure 1 one finds $S\simeq 32$ cm and $D\simeq 30$ cm, and working back to the $(NI)^2/V_r$ value used from his data, $(NI)^2/V_r\simeq 140$. This gives a paraxial focal length (from the 'unified focal length curve') of 17·9 cm. Taking into consideration the separation of principal planes $2(f-z_0)\simeq 9{\cdot}6$ cm, and the correction for spherical aberration ($C_s\simeq 16{\cdot}7$ cm) of $-1{\cdot}8$ cm, both according to figure 5 of this paper, a calculated distance between β-ray source and counter slit of 60·2 cm is obtained. Siegbahn's experimental value is 60·0 cm.

§ 4. A PRACTICAL FOCAL LENGTH CHART

From a practical point of view, it is more convenient to put the data of the lower 'limit-curve' of figure 3, which are representative of the more usual relative lens dimensions, into the form of charts from which the focal length can be directly read off for the various operating conditions. One such chart is shown in figure 4. The straight lines through the origin represent f for the constant value of NI given by the scale on the top and right edges of the chart. The chart has been drawn for $V=V_0=50$ kv ($V_{r0}=53$ kv); its data can be used for other values of V if the NI scale is multiplied by the factor $(V_r/V_{r0})^{1/2}$. Also shown in figure 4 are curves giving the field strength in kilooersted (broken lines), for the values of NI used in the chart, as function of the gap width S_0 in cm (otherwise f and $S+D$ may be measured in any convenient unit of length), to allow the user of the chart a ready estimate of the field strength which will occur in the lens under the operating conditions considered. Obviously, if NI is scaled in the ratio $(V_r/V_{r0})^{1/2}$, the values of H_p written on the broken curves have to be scaled by the same factor.

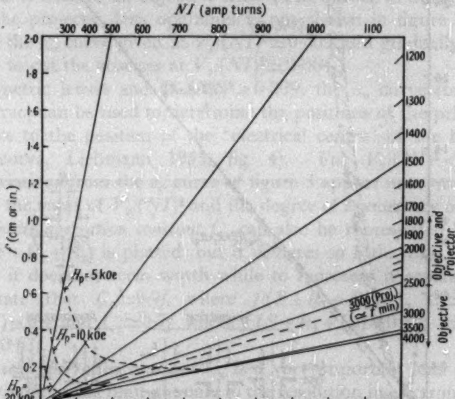


Figure 4. Focal length curves for magnetic lenses of 'standard' geometry. $V=50$ kv, $V_{r0}=53$ kv.

§ 5. ASYMMETRICAL LENSES

In a preceding note (Liebmann 1955) on the axial field distribution in asymmetrical magnetic electron lenses, the two pole-piece bores having the radii R_1 and R_2, it was shown that the field distribution in an asymmetrical lens in the neighbourhood of its maximum is almost identical with that of a symmetrical lens of the same gap width S and a bore diameter $D=R_1+R_2$. As the refracting power of a magnetic lens is mainly determined by the field near its maximum, it follows that the formula (1) can be applied to asymmetrical lenses if D is replaced by R_1+R_2, i.e.

$$f/(S+R_1+R_2)=\alpha(NI)^2/V_r \qquad \ldots\ldots(2)$$

where α has the same value as before. Again, one would expect that lenses in which $0{\cdot}5 \leqslant S/(R_1+R_2) \leqslant 2$ lie close to the lower 'limit-curve' of figure 3, and that the chart, figure 4, can be directly applied if the abscissa $S+D$ is replaced by $S+R_1+R_2$.

To confirm this, the focal lengths of two asymmetrical lenses, with $R_1/R_2=1{\cdot}67$ and $R_1/R_2=2{\cdot}5$, were computed directly for several values of $(NI)^2/V_r$ by tracing electron trajectories through the lens fields, which had been measured by the resistance-network analogue method (Liebmann 1955); the electron trajectories were calculated by the recurrence formulae given by Liebmann (1949). These values of f are indicated by crosses in figure 5, which shows also some of the values computed by Liebmann and Grad (1951) for symmetrical lenses of various S/D ratios. It is seen that the asymmetrical lenses fit in very well with the other types. Similarly, a point evaluated for the 'pinhole' lens, $R_2 \to 0$, would fall exactly on the f curve of figure 5.

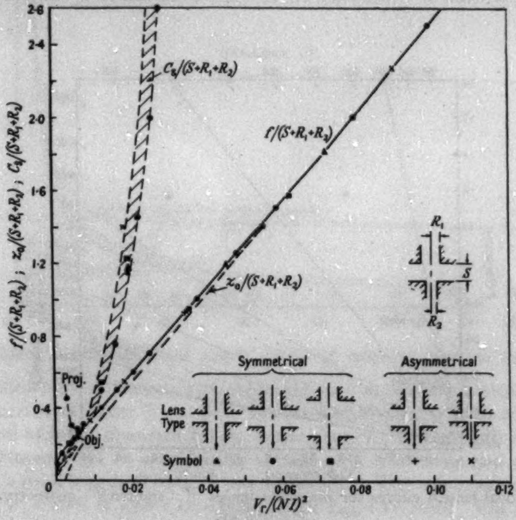


Figure 5.

The focal length f_0 of an asymmetrical lens, and its other lens constants, are independent of whichever side is object or image space for lens excitations $V_r/(NI)^2 \geqslant 0{\cdot}009$. For higher excitations, $V_r/(NI)^2 < 0{\cdot}009$, where the object comes to lie within the lens field, one finds that f_0 and f_1 are given by the curves shown if the object space is on the side of the larger lens bore. If the object space is on the side of the smaller lens bore, the focal length f_0 is slightly greater than is shown by the curve.

§ 6. IMAGE POSITION, CHROMATIC ABERRATION AND SPHERICAL ABERRATION

Apart from the focal length f, the most important optical data required in using an electron lens are the object and image positions z_0 and z_1, the chromatic aberration constant C_c, and the spherical aberration constant C_s (for a definition of C_c and C_s, see Liebmann and Grad 1951).

For a weak lens, the principal planes coincide with the centre plane, and there is no need to consider the image and object positions separately from the focal length. As the lens is made stronger, the principal planes, which are crossed over, move away from the centre planes, and in strong lenses it may be necessary to take account of this fact. As was the case for the focal length f, it is possible to represent the *object position* z_0, for $z_1=\infty$, by a single curve for lenses $0{\cdot}5 \leqslant S/D \leqslant 2$ and approximately for lenses of smaller or larger S/D ratio. This 'unified z_0 curve' is shown by the broken line in figure 5. The distance between the f curves and the z_0 curve gives the *distance of the principal plane* from the centre of the lens (see Liebmann and Grad 1951, fig. 11). Until the object or image has moved into the lens field, i.e. for $V_r/(NI)^2 \geqslant 0{\cdot}009$, image and object positions can be evaluated for any pair of conjugates from f and $f-z_0$ as given in figure 5, whereas for $V_r/(NI)^2 < 0{\cdot}009$ this is only correct for $z_1 \gg z_0$, and if the lens is used as an objective lens. (The branch of the z_0 curve, corresponding to the projector lens operation is not shown in figure 5. It would separate from the z_0 curve given at $V_r/(NI)^2 \simeq 0{\cdot}009$ and gradually curve down more strongly to cut the abscissa at $V_r/(NI)^2 \simeq 0{\cdot}004$.)

For asymmetric lenses and $V_r/(NI)^2 \geqslant 0{\cdot}009$, the z_0 curve (taken together with the f_0 curve) can be used to determine the positions of the principal planes (with reference to the position of the 'electrical centre' of the lens, as given by the z_m curve, Liebmann 1955, fig. 4). For $V_r/(NI)^2 < 0{\cdot}009$, small systematic deviations from the z_0 curve of figure 5 appear in asymmetric lenses, depending on the value of $V_r/(NI)^2$ and the degree of asymmetry of the lens.

The *chromatic aberration constant* C_c can also be represented by one single curve if $C_c/(S+R_1+R_2)$ is plotted, but it deviates so little from the f curve in figure 5, that it does not seem worth while to represent it separately. It will suffice to state that $C_c \simeq 0{\cdot}9f$ where $f/(S+R_1+R_2)=1$, $C_c \simeq 0{\cdot}8f$ where $f/(S+R_1+R_2)=0{\cdot}4$ and $C_c \simeq 0{\cdot}7f$ where $f/(S+R_1+R_2)=0{\cdot}19$, with a limiting value of $C_c \simeq 0{\cdot}63f$.

The *spherical aberration constant* C_s is a very important lens parameter, as spherical aberration sets a limit not only to the resolution in electron microscopes, but also to the resolution, or conversely the line intensity, in β-ray spectrometers, and to the smallness (or conversely the brightness) of the focused spot in a cathode-ray tube. Like the other lens constants, C_s can be represented in the form $C_s/(S+R_1+R_2)$, but not as a single curve embracing the values for almost the whole range of lens geometries used in practice. Nevertheless, it is found that over the range of lens geometries represented by the single f curve, $C_s/(S+R_1+R_2)$ lies in a *band* (shown cross-hatched in figure 5) of such small width that the best and worst values of C_s are within ± 10% of the average value. The lowest values of $C_s/(S+R_1+R_2)$ are found for $R_1 \simeq R_2 \simeq S/2$ (i.e. symmetrical lenses with $S/D \simeq 1$).

For $V_r/(NI)^2 \geqslant 0{\cdot}009$, the spherical aberration is then given by $\delta r_1 = MC_s\theta^3$, where M is the magnification and θ the aperture angle in object space independent of the actual positions of object and image planes. For stronger lens excitations, where the object lies within the lens field, this relation, and the value of C_s given in figure 5, apply only for $M \gg 1$, the lens being assumed to be an objective. In this region the C_s values become more dependent on the geometry (i.e. the S/D ratio, or the degree of asymmetry), and the data of the C_s 'band' are here only approximate. For $V_r/(NI)^2 \rightarrow 0$ and $R_1 \simeq R_2 \simeq S/2$, a limiting value $C_s/(S+R_1+R_2) \simeq 0{\cdot}025$ is attained.

ACKNOWLEDGMENT

The writer wishes to thank Dr. T. E. Allibone, the Director of the Laboratory, for permission to publish this paper.

APPENDIX

THE BEST LENS PROPORTIONS FOR HIGHEST RESOLVING POWER IN ELECTRON MICROSCOPY

The resolving power of the electron microscope as determined by the two unavoidable limits of diffraction and of spherical aberration, in the absence of chromatic errors and (axial) astigmatism, is given by the expression $\delta = B\lambda^{3/4}C_s^{1/4}$ where λ is the electron wavelength and B a constant for which various values have been suggested in the literature, depending to some extent on the definition of resolving power adopted, e.g. Glaser has given two alternative values $B = 0{\cdot}56$ (1943) and $B = 0{\cdot}78$ (1949) whereas Haine (1954) has pointed out that B may be as small as 0·43. The question of the optimum lens dimensions for highest resolving power has been discussed before (Liebmann 1951). It is intended here to give a brief representation of the lens data which allows an easier appraisal of the optimum design conditions.

The 'resolution parameter' $\lambda^{3/4}C_s^{1/4}$, which is *independent* of the value of the constant B, is plotted in figure 6 for several values of H_p in the pole piece gap ($H_p = 4\pi NI/10S$) and an accelerating voltage of 50 kv as a function of gap length S in cm. It is seen from the computed points for the three values of

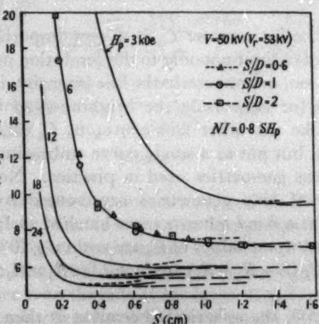


Figure 6. Resolution parameter $\lambda^{3/4} C_s^{1/4}$.

$S/D=0{\cdot}6$, 1 and 2 given for the curve $H_p=6$ kilooersteds but omitted for the other curves, that $\lambda^{3/4}C_s^{1/4}$ is essentially a function of the gap length S, but is over a certain range of S little dependent on the ratio S/D within the S/D range considered (for S/D ratios outside this range, and asymmetrical lenses C_s was seen in §6 to be slightly greater, hence such lens geometries are not optimal for highest resolution). The curves in figure 6 show rather flat optima, indicating that the best theoretical resolving power is not critically dependent on the lens dimensions. For example, while the optimum value of S at 50 kv for $H_p=24$ kOe, the highest practicable value, would be $S=0{\cdot}36$ cm requiring 6700 ampere turns and giving a resolution parameter of $\lambda^{3/4}C_s^{1/4}=5{\cdot}0$ Å, very little resolving power is lost if S is decreased to 0·20 cm or increased to 0·60 cm. Similarly, at the optimum value $S=0{\cdot}36$ cm, the lens bore D could be changed from $D=0{\cdot}20$ cm to $D=0{\cdot}60$ cm without effect on the resolving power, allowing a range of focal lengths from 0·19 cm to 0·44 cm. A still greater latitude in 'nearly optimum' electron lens design would result if a small reduction in the best possible resolving power were permitted. The curves of figure 6 can be scaled to serve for other values of V_r if the values of S (and consequently of NI) are multiplied by $(V_r/V_{r0})^{1/2}$ and the values of the resolution parameter $\lambda^{3/4}C_s^{1/4}$ by $(V_r/V_{r0})^{-1/4}$, where $V_{r0}=53$ kv is the value of V_r used in figure 6.

REFERENCES

BECKER, H., and WALLRAFF, A., 1938, *Arch. Elektrotech.*, **32,** 644.

GLASER, W., 1943, *Z. Phys.*, **121,** 647; 1949, *Acta Phys. Austriaca*, **3,** 38.

HAINE, M. E., 1954, *Advances in Electronics and Electron Physics*, Ch. 6 (New York: Academic Press Inc.), p. 298.

LENZ, F., 1950, *Z. Angew. Phys.*, **2,** 448.

LIEBMANN, G., 1949, *Proc. Phys. Soc.* B, **62,** 753; 1951, *Ibid.*, **64,** 972; 1955, *Ibid.*, **68,** 679.

LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64,** 956.

LYLE, T. R., 1902, *Phil. Mag.*, **3,** 310.

VAN MENTS, M., and LE POOLE, J. B., 1947, *Appl. Sci. Res.* B, **1,** 3.

RAMBERG, E. G., 1942, *J. Appl. Phys.*, **13,** 582.

SIEGBAHN, K., 1944, *Ark. Mat. Astr. Fys.* A, **30,** No. 1.

Field Emission from Semiconductors

By R. STRATTON

Research Department, Metropolitan-Vickers Electrical Co. Ltd., Trafford Park, Manchester

MS. received 30th December 1954 *and in modified form* 21*st April* 1955

Abstract. The field emission current from a simple semiconductor has been evaluated in a manner similar to that employed by Fowler and Nordheim for a metal. It is shown that special effects will be associated with the state of the semiconductor surface. The presence of an appreciable number of surface states should give very low currents until fields high enough to break down the internal barrier are applied. When this occurs, or if there are no surface states at all, the field will penetrate into the semiconductor and lead to a lowering of the conduction band edge. This causes degeneracy in the electron distribution near the surface region and a greatly enhanced emission not unlike that of a metal.

It is pointed out that the current–field emission characteristic dominated by surface states should be strongly temperature dependent while that dominated by field penetration is not.

The limits of applicability for the various results are indicated.

---

§ 1. Introduction

Although a great deal of attention has been paid to the theory of the field emission from a metal no detailed analysis has been given for similar effects in semiconductors. To a large extent the theory for the two cases can be given a parallel development since firstly, the transmission coefficients for the surface barriers are similar and, secondly, the statistics for the current carriers in the semiconductor (Maxwell) are a limiting case of those of the metal (Fermi–Dirac). The theory below extends the original Fowler–Nordheim theory to include semiconductors. It will however be seen that special effects are associated with semiconductors due to the state of their surface and the fact that an external field applied to a dielectric has an appreciable penetration into the interior.

Figure 1 is a potential energy diagram for the electrons in an n-type semiconductor. The Fermi level F is below the bottom of the conduction band C, the Fermi energy ζ being negative and usually less than 1 ev. The work function, denoted by χ, is the energy required to remove an electron from the Fermi level and place it just outside the metal. If ψ is the potential energy of electrons in the conduction band

$$\psi=\chi+\zeta. \qquad \ldots\ldots(1)$$

It is supposed that the forbidden gap width E_g is much larger than $|\zeta|$ so that no minority carriers (holes in the full band) need be considered. Finally we consider an electric force F perpendicular to the surface of the semiconductor but neglect any image force or penetration effects. These will be taken into account later on.

Take the x axis perpendicular to the surface and consider the electrons moving towards the surface. Let

$$E_x = p_x^2/2m \qquad \ldots\ldots(2)$$

where p_x is the momentum of an electron in the x direction and m is its effective mass. The electron will have a probability $D(E_x)$ of tunnelling through the barrier which is a function of E_x, F and the constants of the solid. (The components p_y and p_z are conserved and do not effect $D(E_x)$.) If we put $n_x(E_x)dE_x$ equal to the total number of electrons striking unit area of the surface in unit time and having E_x values in the range E_x, E_x+dE_x, then the emitted current will be

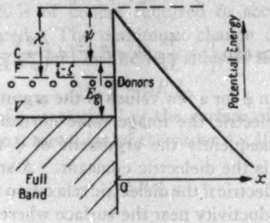


Figure 1. Semiconductor.

$$-I = -e\int_0^\infty n_x(E_x)D(E_x)\,dE_x \qquad \ldots\ldots(3)$$

(e is the positive electronic charge).

§ 2. TRANSMISSION COEFFICIENTS

Fowler and Nordheim (1928) considered the barrier shown in figure 2 in their calculation for the field emission from a metal; i.e. if V is the electronic potential energy and F is the applied field measured in ev cm^{-1} (gradient of the potential energy) then

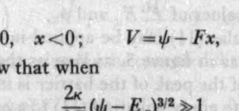


Figure 2. Surface barrier.

$$V=0,\quad x<0;\qquad V=\psi - Fx,\quad x>0. \qquad \ldots\ldots(4)$$

Their calculations show that when

$$\frac{2\kappa}{3F}(\psi - E_x)^{3/2} \gg 1 \qquad \ldots\ldots(5)$$

where

$$\kappa = \left(\frac{8m\pi^2}{h^2}\right)^{1/2} \simeq 5{\cdot}16\times 10^{-7}(\text{ev})^{-1/2}\,\text{cm}^{-1} \qquad \ldots\ldots(6)$$

then

$$D(E_x) \sim \frac{4E_x^{1/2}(\psi - E_x)^{1/2}}{\psi}\exp\left[-\frac{4\kappa}{3F}(\psi - E_x)^{3/2}\right]. \qquad \ldots\ldots(7)$$

Condition (5) is necessary because Bessel functions have been replaced by their asymptotic expressions in equation (7).

Nordheim (1928) also considered the effect of the image force on the electron outside the metal, when its potential energy becomes (broken line in figure 2)

$$V=0, \quad x<x_0; \qquad V=\psi-Fx-e^2/4x, \quad x>x_0 \qquad \ldots\ldots(8)$$

x_0 being given by

$$\psi=Fx_0+e^2/4x_0. \qquad \ldots\ldots(9)$$

He found

$$D(E_x)\simeq \exp\left\{-\frac{4\kappa}{3F}(\psi-E_x)^{3/2}\phi\left(\frac{e\sqrt{F}}{\psi-E_x}\right)\right\} \qquad \ldots\ldots(10)$$

and computed the function ϕ for a few values of the argument.

For the case of a dielectric the image force correction is to be multiplied by $(\epsilon-1)/(\epsilon+1)$ and consequently the argument of ϕ is to be multiplied by $[(\epsilon-1)/(\epsilon+1)]^{1/2}$ where ϵ is the dielectric constant. A semiconductor would be expected to behave as a dielectric if the dielectric relaxation time $\epsilon/4\pi\sigma$ is sufficiently large. (Here σ is the conductivity near the surface where charges are displaced.) This case will be assumed in the rest of the paper since in any case the factor is not very different from unity for the ϵ values of typical semiconductors.

Equation (10) is based on the general result (see Mott and Sneddon 1948) that the transmission coefficient for a 'smooth' barrier such as is shown in figure 3 is

$$D(E_x)=\exp\left[-2\int_{x_1}^{x_2}\kappa[V(x)-E_x]^{1/2}\,dx\right]. \qquad \ldots\ldots(11)$$

This result is only valid when the barrier is 'smooth' enough to fulfil the conditions stated in the Appendix which, as there shown, may be reduced to the condition that the integral in the exponent of equation (11) is large compared with unity. This incidentally is similar to condition (5).

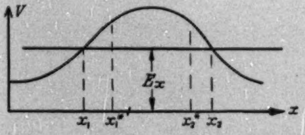


Figure 3. 'Smooth' barrier (cf. Appendix).

Numerical calculation shows that the image force barrier obeys the required conditions for reasonable values of F, E_x and ψ.

Exactly the same formulae (11) may be applied in the case where there is an additional internal barrier as in figure 5, as long as the same conditions are still obeyed. Only the shape of the peak of the barrier is important.

It is however incorrect to apply equation (11) to a barrier with discontinuities at the peak as in figure 1, or to the simple barrier between two metals in contact, as has sometimes been done in the literature.

Finally we may remark that the implied upper limits on F do not restrict the applicability of the results since the currents obtained for the high fields, which would occur if F exceeded these limits, would disrupt the material.

§ 3. Field Penetration

When a field is applied to the surface of a metal or a dielectric it induces a surface charge which may be calculated by ordinary electrostatic theory. Fan (1942) has examined the distribution of this surface charge for the case of a metal. It is found that the lower edge of the conduction band C dips down as the surface is approached from the interior, i.e. there is a decrease in potential energy. Since the Fermi level is constant this implies an increase in the (positive) Fermi energy near the surface which is of course required to accommodate the additional electronic charge induced. The maximum change ΔR in the Fermi energy, calculated by equating the field on the two sides of the barrier, turns out to be

$$F=\Delta R/l \qquad \ldots\ldots(12)$$

where l is the 'depth of penetration' of the surface charge (exponential decay). Theory indicates that in a metal l is of the order 3×10^{-9} cm so that even for the highest fields

$$\Delta R \ll \chi \qquad \ldots\ldots(13)$$

and may be neglected.

In the case of semiconductors the position is different due to the relatively low density of carriers. This has been pointed out by Margulis (1947) who derived the potential drop when the field is so low that the carriers in the barrier are always non-degenerate. We will however be primarily interested in the case of stronger fields which cause the carriers to become degenerate. Let U(ev) be the decrease in potential energy of any point a distance x from the surface, then (cf. figure 4)

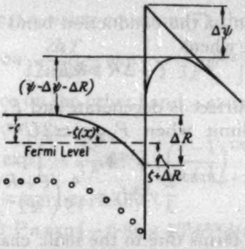


Figure 4. Field penetration into a semiconductor (degenerate case).

$$\left.\begin{aligned} &U(0)=\Delta R; \qquad U(x)\to0 \quad\text{as}\quad x\to\infty \\ \text{and}\quad &\left(\frac{dU}{dx}\right)_{x=0}=\frac{F}{\epsilon}; \qquad \frac{dU}{ax}\to0 \quad\text{as}\quad x\to0. \end{aligned}\right\} \qquad \ldots\ldots(14)$$

In general Poisson's equation for the potential distribution may now be written as

$$\frac{d^2U}{dx^2}=-\frac{4\pi ne^2}{\epsilon}\left[\exp\left(-\frac{U}{kT}\right)-\frac{4}{\sqrt{\pi}}\frac{N_0}{n}F_{1/2}\left(\frac{U+\zeta}{kT}\right)\right] \qquad \ldots\ldots(15)$$

where

$$N_0=(2\pi mkT/h^3)^{3/2} \qquad \ldots\ldots(16)$$

and

$$F_c(y)=\int_0^\infty \frac{x^c dx}{1+\exp(x-y)}. \qquad \ldots\ldots(17)$$

The first term in equation (15) is due to the positive ionic charges while the latter is due to the electronic charges.

Integrating once, using the boundary conditions (14) gives

$$F^2 = 2\left(\frac{\epsilon \boldsymbol{k}T}{x_0}\right)^2 \left[-1 + \exp\left(-\frac{\Delta R}{\boldsymbol{k}T}\right) + \left(\frac{x_0}{X_0}\right)^2 \frac{2}{\sqrt{\pi}} \int_{\zeta/\boldsymbol{k}T}^{(\zeta+\Delta R)/\boldsymbol{k}T} F_{1/2}(y)\, dy\right] \qquad \ldots\ldots(18)$$

where

$$x_0 = (\epsilon \boldsymbol{k}T/8\pi n \boldsymbol{e}^2)^{1/2} = X_0(2N_0/n)^{1/2}. \qquad \ldots\ldots(19)$$

Now when classical statistics hold right up to the surface, i.e. when

$$-(\Delta R + \zeta) \gg \boldsymbol{k}T \qquad \ldots\ldots(20)$$

equation (18) reduces to

$$\sinh \Delta R/2\boldsymbol{k}T = F x_0/2\epsilon \boldsymbol{k}T \qquad \ldots\ldots(21)$$

this being the case treated by Margulis in considering thermionic emission.

Substituting from equation (21) it can be shown that condition (20) is equivalent to

$$F \ll \epsilon \boldsymbol{k}T/X_0 \qquad \ldots\ldots(22)$$

i.e. at room temperature

$$F \ll \epsilon^{1/2} 1{\cdot}5 \times 10^6 \text{ ev cm}^{-1}. \qquad \ldots\ldots(22a)$$

If however the bottom of the conduction band dips well below the Fermi level near the surface, i.e. when

$$\zeta + \Delta R \gg \boldsymbol{k}T \qquad \ldots\ldots(23)$$

the electron gas near the surface is degenerate and $F_{1/2}(y)$ is very small except for values near the upper limit when $F_{1/2}(y) \simeq 2U^{3/2}/3$. Hence the integral is approximately

$$\int_0^{(\zeta+\Delta R)/\boldsymbol{k}T} \frac{2}{3} y^{3/2} dy = \frac{4}{15}\left(\frac{\zeta + R}{\boldsymbol{k}T}\right)^{3/2}$$

and (neglecting the small terms due to the ionic charge)

$$\zeta + \Delta R \simeq \nu F^{4/5} \qquad \ldots\ldots(24)$$

where

$$\nu = \left(\frac{15\sqrt{\pi}}{16}\right)^{2/5} \boldsymbol{k}T \left(\frac{X_0}{\epsilon \boldsymbol{k}T}\right)^{4/5}. \qquad \ldots\ldots(25)$$

The constant ν is actually independent of T ($X_0 \propto T^{-1/4}$) and is equal to

$$\nu = \frac{1}{\epsilon^{2/5}}\left[\frac{225}{\pi^4 2^{11}} \frac{\boldsymbol{h}^6}{\boldsymbol{m}^3 \boldsymbol{e}^4}\right]^{1/5} \simeq \epsilon^{-2/5} 3{\cdot}568 \times 10^{-7} \text{ ev}^{1/5} \text{ cm}^{4/5}. \qquad \ldots\ldots(25a)$$

Condition (23) is essentially (cf. equation (24))

$$F \gg \left(\frac{16}{15\sqrt{\pi}}\right)^{1/2} \frac{\epsilon \boldsymbol{k}T}{X_0} \simeq 0{\cdot}78 \frac{\epsilon \boldsymbol{k}T}{X_0} \text{ ev cm}^{-1} \qquad \ldots\ldots(26)$$

i.e. the converse of condition (22), and at room temperature it reduces to

$$F \gg \epsilon^{1/2} 1{\cdot}2 \times 10^6 \text{ ev cm}^{-1}. \qquad \ldots\ldots(26a)$$

When this condition on F applies the Fermi energy near the surface is positive and greater than $\boldsymbol{k}T$.

§ 4. Emission from Semiconductor without Surface Barrier

If Maxwellian statistics can be applied right up to the surface of the semiconductor

$$n_x(E_x) = en\left(\frac{kT}{2\pi m}\right)^{1/2} \frac{\exp(-E_x/kT)}{kT}$$

and this can be substituted in equation (3). Ignoring field penetration for the moment we can also substitute equation (7) for $D(E_x)$ if image forces are not considered, or equation (10) if they are.

In either case the exponential can be simplified if

$$2\kappa\psi^{1/2}/F \gg 1/kT \qquad \ldots\ldots(27)$$

which implies that $D(E_x)$ increases less rapidly with E_x than $n_x(E_x)$ decreases. This is simply the condition that emission by tunnelling is much more important than thermionic emission; if the condition were reversed (smaller F or larger T) thermionic emission would become important and the integrand would diverge. Condition (27) and condition (5) together limit F to the finite range $2\kappa\psi^{1/2}kT \ll F \ll 2\kappa\psi^{3/2}/3$ which expressed numerically for room temperature means $2{\cdot}6 \times 10^6\psi^{1/2} \ll F \ll 3{\cdot}4 \times 10^7\psi^{3/2}$ (ψ in ev, F in ev cm^{-1}. The degree of inequality implied by (27) need not be large, since the quantities occur as exponents.)

Ignoring the image force, i.e. using equation (7), the result† becomes

$$I \simeq en\frac{2kT}{(2m\psi)^{1/2}}\exp\left\{-\frac{4\kappa}{3F}\psi^{3/2}\right\}. \qquad \ldots\ldots(28)$$

Including the image force on the other hand, i.e. using equation (10), leads to the result

$$I = ne\left(\frac{kT}{2\pi m}\right)^{1/2}\exp\left[-\frac{4\kappa}{3F}\psi^{3/2}\phi\left\{\left(\frac{\epsilon-1}{\epsilon+1}\right)^{1/2}\frac{e\sqrt{F}}{\psi}\right\}\right]. \qquad \ldots\ldots(29)$$

At room temperature this is numerically

$$I = 4{\cdot}25 \times 10^{-13} n \exp[-6{\cdot}86 \times 10^7 \psi^{3/2}\phi(u)/F] \qquad \ldots\ldots(29a)$$

with

$$u = 3{\cdot}62 \times 10^{-4}\left(\frac{\epsilon-1}{\epsilon+1}\right)^{1/2} F^{1/2}/\psi$$

I being in amp cm^{-2}, ψ in ev and F in ev cm^{-1}. This result as it stands may not be applied to real semiconductors because of the effect of field penetration discussed in the previous section.

If the lowering of the conduction band is not great enough to invalidate Maxwellian statistics near the surface, the value of n requires correction by a factor $\exp(\Delta R/kT)$ where ΔR is given by equation (21).

If on the contrary the effect of the external field is to produce a region near the surface where degeneracy prevails (as seems likely from inequality (26*a*)) $n_x(E)$ will have a value appropriate to a metal. The calculation for the field emission current is parallel to that for a metal (Fowler and Nordheim 1928) where the work function is $\psi - (\Delta R + \zeta) = \psi - \nu F^{4/5}$ and the Fermi energy is $\Delta R + \zeta = \nu F^{4/5}$. However now the Fermi energy may be less than the work function by an order of magnitude so that some of the approximations cannot be made. Substituting into equation (3) and following Fowler and Nordheim

---

† The result $I = en(kT/2\pi m)^{1/2}\exp(-4\kappa\psi^{3/2}/3F)$ which is sometimes quoted in the literature is based on the assumption that one can use expression (11) for $D(E_x)$ when the barrier has a sharp peak. We have already discussed why this is inconsistent.

$$I=\frac{\boldsymbol{e}}{8\pi \boldsymbol{h}}\frac{F^2}{\psi-\nu F^{4/5}}\exp\left[-\frac{4}{3}\frac{\kappa(\psi-\nu F^{4/5})^{3/2}}{F}\right]\phi(u)$$

$$\times\left[1-\left(1+\frac{2\kappa\nu F^{4/5}(\psi-\nu F^{4/5})^{1/2}}{F}\right)\exp\left\{\frac{2\kappa\nu F^{4/5}(\psi-\nu F^{4/5})^{1/2}}{F}\right\}\right] \qquad \ldots\ldots(30)$$

where

$$u=\left(\frac{\epsilon-1}{\epsilon+1}\right)^{1/2}\frac{\boldsymbol{e}\sqrt{F}}{\psi-\nu F^{4/5}} \quad \text{and} \quad F\ll \tfrac{2}{3}\kappa(\psi-\nu F^{4/5})^{3/2}\phi \qquad \ldots\ldots(31)$$

for the case where image forces are included. If

$$\Delta R+\zeta=\nu F^{4/5}<\psi \qquad \ldots\ldots(32)$$

these formulae reduce to

$$I\simeq\frac{\boldsymbol{e}}{8\pi \boldsymbol{h}}\frac{F^2}{\psi}\exp\left(-\frac{4}{3}\frac{\kappa\psi^{3/2}\phi(u)}{F}\right)\left[\exp\left\{\frac{2\kappa\nu\psi^{1/2}}{F^{1/5}}\right\}-\left(1+\frac{2\kappa\nu\psi^{1/2}}{F^{1/5}}\right)\right] \qquad \ldots\ldots(30a)$$

where

$$u\simeq\left(\frac{\epsilon-1}{\epsilon+1}\right)^{1/2}\frac{\boldsymbol{e}\sqrt{F}}{\psi} \quad \text{and} \quad F\ll \tfrac{2}{3}\kappa\psi^{3/2}\phi. \qquad \ldots\ldots(31a)$$

These formulae (30*a*) and (31*a*) would have been obtained if we had expanded $D(E_x)$ about the bottom of the conduction band rather than the Fermi level as is done by Fowler and Nordheim. In fact, condition (32) is really assumed when we use the expanded form of $D(E_x)$ for the whole range of electron energies.

Numerically the results become

$$I=1{\cdot}55\times 10^{-6}\frac{F^2}{\psi}\exp\left[-6{\cdot}86\times 10^7\psi^{3/2}\phi(u)/F\right]$$

$$\times\left[\exp\left\{36{\cdot}8\frac{\psi^{1/2}}{\epsilon^{2/5}F^{1/5}}\right\}-\left(1+36{\cdot}8\frac{\psi^{1/2}}{\epsilon^{2/5}F^{1/5}}\right)\right]\text{amp cm}^{-2} \qquad \ldots\ldots(30b)$$

where

$$u=3{\cdot}62\times 10^{-4}\left[\frac{\epsilon-1}{\epsilon+1}\right]^{1/2}\frac{F^{1/2}}{\psi} \qquad \ldots\ldots(31b)$$

and $1{\cdot}2\times 10^6\epsilon^{1/2}\ll F\ll 3{\cdot}4\times 10^7\psi^{3/2}\phi(u)$.

4.1. *Constancy of the Fermi Level near the Surface*

In the previous section we have assumed that the Fermi level is constant throughout the semiconductor. This implies thermal equilibrium, i.e. no resultant current flows in the semiconductor. Thus in the surface region the diffusion current due to non-uniform carrier density is balanced by the conduction current. If the diffusion current and the conduction current are only slightly different, i.e. if the current I drawn out is much smaller than either, we can still neglect the very small change of Fermi level.

If $\zeta(x)$ is the Fermi energy measured from the bottom of the conduction band, we may write the current in the semiconductor as

$$I=\frac{\sigma}{\boldsymbol{e}}\left(-\frac{\partial U}{\partial x}+\frac{\partial\zeta(x)}{\partial x}\right) \qquad \ldots\ldots(33)$$

and since $\partial U/\partial x$ will be of the order of 10^7 ev cm^{-1} the resistive voltage gradient I/σ within the semiconductor would have to be improbably large to affect the Fermi level.

§ 5. EMISSION FROM SEMICONDUCTOR WITH SURFACE BARRIER DUE TO SURFACE STATES

Many of the contact phenomena of semiconductors can be explained by assuming that states having energies within the forbidden band exist on the surface. These states give rise to an internal potential barrier (e.g. Schottky exhaustion layer) at the free surface of a semiconductor which, if the density of the states is high enough, may be almost unaffected by putting the surface in contact with a metal. We consider then the model of figure 5 where D represents donor-type states (energy E_d below the conduction band) and A acceptor states (energy E_a below the conduction band). Thus the D states are neutral when occupied by electrons and the A states when occupied by holes (of electrons). This model with

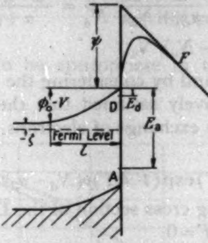


Figure 5. Internal barrier due to surface states.

$$\exp\left(\frac{E_d}{kT}\right),\ \exp\left(\frac{E_g-E_a}{kT}\right) \ll \exp\left(\frac{E_g}{kT}\right) \qquad \ldots\ldots(34)$$

(E_g = gap width) has been used by Brattain and Bardeen (1953) to describe the surface properties of Ge. For simplicity it will be assumed that the surface states are sufficiently numerous to provide all the charge induced by the external field i.e. that field penetration does not occur.

We now introduce the following symbols: ϕ_0(ev) = height of internal barrier when $F=0$; (ϕ_0-V) (ev) = height of internal barrier when $F\neq 0$; N_d, N_a = number of donor (acceptor) surface states per cm^2; n_{d0}, p_{a0} = number of electrons on donor surface states, holes on acceptor surface states when $F=0$; n_d, p_a = number of electrons on donor surface states, holes on acceptor surface states when $F\neq 0$; E_0/ϵ (ev cm^{-1}) = field at the top of the barrier inside the semiconductor when $F=0$; E/ϵ (ev cm^{-1}) = field at the top of the barrier inside the semiconductor when $F\neq 0$.

The charge on the surface states is determined by

$$\frac{F+E}{4\pi e^2} = -(N_d-n_d)+(N_a-p_a) \qquad \ldots\ldots(35)$$

which reduces to

$$\frac{E_0}{4\pi e^2} = -(N_d-n_{d0})+(N_a-p_{a0}) \qquad \ldots\ldots(36)$$

when $F=0$.

It should be observed that this assumes a high density of surface states. Thus from equation (35) $N_a > (E+F)/4\pi e^2 \simeq 10^{13}$ cm^{-2} when $(E+F)\simeq 2\times 10^7$ ev cm^{-1}

and further (36) can only be satisfied if N_d and N_a are of equal orders of magnitude since E_0 is far smaller than F. For smaller surface densities there may be a potential drop inside the semiconductor as in § 3 due to the inability of the surface states to balance the applied field F.

It is easily shown (and is almost obvious from examination of figure 5) that if

$$\exp(E_g/2\boldsymbol{k}T) \gg \exp\{(\phi_0-\zeta)/\boldsymbol{k}T\} \qquad \ldots\ldots(37)$$

then the acceptor states will all carry electrons i.e. $p_{a0}=p_a=0$, thus

$$\frac{F+E}{E_0}=\frac{n_d+N_a-N_d}{n_{d0}+N_a-N_d}=\frac{n_d/N_d+\beta}{\alpha+\beta} \qquad \ldots\ldots(38)$$

where $\alpha=n_{d0}/N_d$ and $\beta=(N_a-N_d)/N_d$.

The value of n_d can be found by considering the dynamic equilibrium of the donor traps D. (It is effectively assumed that the acceptor traps are always full.) If we only consider the exchange of electrons with the conduction band†, then

$$\boldsymbol{e}n(\boldsymbol{k}T/2\pi\boldsymbol{m})^{1/2}\exp(-\phi_0/\boldsymbol{k}T)\exp(V/\boldsymbol{k}T)p(N_d-n_d)=\boldsymbol{e}n_d/\tau \qquad \ldots\ldots(39)$$

where p is the average trapping cross section of the D states and τ is the lifetime of electrons in them. When $F=0$

$$\boldsymbol{e}n(\boldsymbol{k}T/2\pi\boldsymbol{m})^{1/2}\exp(-\phi_0/\boldsymbol{k}T)p(N_d-n_{d0})=\boldsymbol{e}n_{d0}/\tau. \qquad \ldots\ldots(40)$$

Thus dividing (39) by (40)

$$\exp\left(\frac{V}{\boldsymbol{k}T}\right)=\frac{n_d}{n_{d0}}\,\frac{N_d-n_{d0}}{N_d-n_d}=\frac{n_d}{n_{d0}}\,\frac{1-\alpha}{1-\alpha n_d/n_{d0}}$$

whence

$$\frac{n_d}{N_d}=\left[1+\frac{1-\alpha}{\alpha}\exp\left(-\frac{V}{\boldsymbol{k}T}\right)\right]^{-1}. \qquad \ldots\ldots(41)$$

Finally eliminating the ratio n_d/N_d from (38) and (41) we find

$$\frac{\alpha+\beta}{E_0}(F+E)=\left[1+\frac{1-\alpha}{\alpha}\exp\left(-\frac{V}{\boldsymbol{k}T}\right)\right]^{-1}+\beta. \qquad \ldots\ldots(42)$$

There is also a relation between E and V which depends on the model used for the barrier. The (algebraically) simplest case is that of a 'Mott' barrier where the field is uniform across most of the barrier (impurity free layer) and charges enter a small region. Physically a 'Schottky' space charge barrier is more plausible but the effect of ionic charge density on ΔR is negligible.

The emitted current will be given by equation (3) with

$$n_x(E_x)=n\boldsymbol{e}\left(\frac{\boldsymbol{k}T}{2\pi\boldsymbol{m}}\right)^{1/2}\exp\left(\frac{V-\phi_0}{\boldsymbol{k}T}\right)=I_0\exp\frac{V}{\boldsymbol{k}T}$$

(where I_0 is appropriately defined) and with $D(E_x)$ given by equation (10) which allows for the image force. As in deriving equation (29) we reach the result

$$I=I_0\exp\frac{V}{\boldsymbol{k}T}\exp\left[-\frac{4}{3}\frac{\kappa}{F}\psi^{3/2}\phi\left\{\left(\frac{\epsilon-1}{\epsilon+1}\right)\right\}^{1/2}\frac{\boldsymbol{e}\sqrt{F}}{\psi}\right]. \qquad \ldots\ldots(43)$$

(For numerical values see equation (29*a*).)

---

† Neglecting the exchange of electrons on the D states with the valence band, i.e. hole emission and capture, can easily be justified by methods analogous to those used by Shockley and Read (1952) provided that the capture cross sections for holes and electrons are comparable and conditions (34) and (37) hold.

§ 6. RESULTS AND DISCUSSIONS

The variation of the emission current as a function of the applied field is plotted in figure 6 for the various cases dealt with.† The graphs continue as approximately straight lines beyond the right-hand edge of the figure. Typical values of the constants have been assumed and are as follows:

$T = 293°\text{K}$
$\psi = 4$ ev
$\epsilon = 7$
$\phi = 0{\cdot}5$ev

$E_0 = 4{\cdot}4 \times 10^6$ ev cm^{-1}
$\alpha = 7 \times 10^{-3}$
$\beta = 0{\cdot}2$
$n = 10^{18}$ cm^{-3}

The values are likely to be appropriate in the case of SiC and lead to $N_d \simeq 1{\cdot}2 \times 10^{13}$ cm^{-2}.

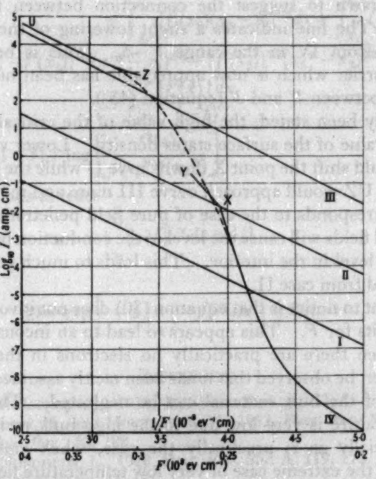


Figure 6. Field emission currents from a semiconductor plotted as functions of $1/F$. I, no image force (equation (28)); II, image force (equation (29)); III, image force and field penetration (equation (30*a*)); IV, image force and surface states (equation (43)).

Curves for cases I and II of figure 6 are very similar to those for a metal, the factor F^2 in front of the exponential in the metal case being relatively unimportant. The semiconductor currents are however strongly temperature dependent due to the factor n. It has already been pointed out that neither curve I nor even curve II can apply to a real semiconductor since variations in the lower edge of the conduction band near the surface must be considered. If there is a sufficiently high density of surface states, the internal barrier produced will cause a drop in current with respect to II for low fields (cf. curve IV). However as the field increases the barrier is lowered and the current begins to increase much more rapidly until curve IV intersects curve II at X which corresponds to the complete breakdown of the barrier ($V=\phi_0$, when $F_0 \simeq 2{\cdot}5 \times 10^7$ ev cm^{-1}).

---

† The values of the Fowler–Nordheim function ϕ have been taken from a recently published table by Burgess *et al.* (1953).

For fields higher than this critical value (say F_0) field penetration will occur and the emitted current is governed by an equation of type (30) (case III), except that the field used to determine ΔR is the applied field F reduced by that due to the now almost completely filled surface states, i.e. F_0. When $F-F_0$ is sufficiently small ($<1{\cdot}5 \times 10^6$ ev cm^{-1}), ΔR is given by equation (21), while the current is derived from equation (29) where n is increased by the factor $\exp(\Delta R/\boldsymbol{k}T)$. This corresponds to portion XY of curve IV. If however, $F-F_0$ is sufficiently large ($>3{\cdot}2 \times 10^6$ ev cm^{-1}), equation (30) must be used with F replaced by $F-F_0$ in the last factor only. (The other factors correspond to the penetration coefficient.) This corresponds to the portion UZ of curve IV. Finally, a broken line has been drawn to suggest the connection between the two regions of approximation. The line indicates a slight lowering of the current below the value calculated from IV in the range $V \rightarrow \phi_0$. This is because the electron density in the barrier which is now appreciable has been neglected in deriving the relationship between V and F (equation (43)).

As has already been stated, the high value of the critical field F_0 is in turn due to the high value of the surface states density. Lower values of the surface states density would shift the point X down curve II while the portion of the curve corresponding to UZ would approach curve III more and more closely.

Curve III corresponds to the case of pure field penetration when no surface states exist and all fields will cause the level of the conduction band near the surface to drop below its level in the interior. This leads to much greater field emission than that obtained from case II.

A further point to notice is that equation (30) does not involve the temperature except in the limits for F. This appears to lead to an inconsistency at very low temperatures when there are practically no electrons in the conduction band. It should, however, be observed that it has been tacitly assumed that the resistance to current flow of the bulk material can be neglected. This will not be true when the temperature is very low due to the high bulk resistance of the semiconductor, especially as is practically the case, when emission occurs from a small area. In the extreme case of very low temperature field emission may be accompanied by internal breakdown. It is the temperature dependence which appears to be the most important criterion for determining the nature of the semiconductor surface.

ACKNOWLEDGMENTS

The author wishes to thank Professor H. Fröhlich for advice on this work, Dr. R. W. Sillars for suggesting the problem and the interest he has taken in the work at all stages, Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd, for permission to publish this paper.

APPENDIX

THE RANGE OF VALIDITY OF EQUATION (11)

It can be shown (Jeffreys 1942) that the derivation of equation (11) for $D(E_x)$ implies that points x_1^* and x_2^* (cf. figure 3) exist inside the peak such that

$$\left|\frac{f'(x_1)}{f''(x_1)}\right| \gg [x_1^* - x_1] \gg \left|\frac{1}{f'(x_1)}\right|^{1/3}; \qquad |f'(x_1^*)| \ll |f(x_1^*)|^{3/2}$$

and

$$\left|\frac{f'(x_2)}{f''(x_2)}\right| \gg [x_2 - x_2^*] \gg \left|\frac{1}{f'(x_2)}\right|; \qquad |f'(x_2^*)| \ll |f(x_2^*)|^{3/2} \qquad \ldots\ldots\text{(A1)}$$

where $f(x) = \kappa^2[E_x - V(x)]$ and $x_2^* - x_1^* \geqslant 0$. Then

$$\begin{aligned}\int_{x_1}^{x_2} (-f)^{1/2} dx &\simeq \int_{x_1}^{x_1^*} [-f'(x_1)(x-x_1)]^{1/2}\, dx \\ &\qquad + \int_{x_2^*}^{x_2} [f'(x_2)(x_2 - x)]^{1/2} dx + \int_{x_1^*}^{x_2^*} (-f)^{3/2} \frac{dx}{(-f)} \\ &\geqslant \tfrac{2}{3}[\{-f'(x_1)\}^{1/2}(x_1^* - x_1)^{3/2} + \{f'(x_2)\}^{1/2}(x_2^* - x_2)^{3/2}] \\ &\qquad + \int_{x_1^*}^{x_2^*} \frac{|f'(x)|\, dx}{(-f)} \\ &\geqslant \tfrac{2}{3}[1+1] + \log\left[\frac{f^2(z)}{f(x_1^*)f(x_2^*)}\right] \qquad \ldots\ldots\text{(A2)}\end{aligned}$$

where $f'(z) = 0$. Thus at least

$$\int_{x_1}^{x_2} [-f]^{1/2}\, dx \geqslant 4/3 \qquad \ldots\ldots\text{(A3)}$$

and

$$D = \exp\left[-2\int_{x_1}^{x_2} [-f]^{1/2}\, dx\right] \leqslant 7\%.$$

REFERENCES

BRATTAIN, W. H., and BARDEEN, J., 1951, *Bell Syst. Tech. J.*, **32**, 1.

BURGESS, R. E., KROEMER, H., and HOUSTON, J. M., 1953, *Phys. Rev.*, **90**, 515.

FAN, H. Y., 1942, *Phys. Rev.*, **62**, 388.

FOWLER, R. H., and NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **119**, 173.

JEFFREYS, H. J., 1942, *Phil. Mag.*, **33**, 451.

MARGULIS, N., 1947, *J. Phys. U.S.S.R.*, **11**, 67.

MOTT, F. N., and SNEDDON, J. N., 1948, *Wave Mechanics and its Applications* (Oxford: Clarendon Press).

NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **121**, 626.

SHOCKLEY, W., and READ, W. T., JR., 1952, *Phys. Rev.*, **87**, 825.

The Absorption Spectrum of Excited Crystals of Cadmium Sulphide

By A. HALPERIN† and G. F. J. GARLICK

Department of Physics, University of Birmingham

MS. received 9*th May* 1955 *and in final form* 19*th May* 1955

Abstract. Crystals of cadmium sulphide show new absorption bands when excited by visible light (0·546 μ). The bands extend from the excitation region out to 1·4 μ with a maximum intensity at 0·78 μ. The bands are stable in the dark at low temperatures but are thermally bleached on warming. They are also bleached by radiation of their own wavelength. Correlation is found between thermal bleaching characteristics and electrical conductivity characteristics. A model is proposed in which thermal bleaching and electrical conduction on warming result from the escape of trapped electrons while optical bleaching frees positive holes which migrate to traps and recombine with trapped electrons.

---

§ 1. Introduction

Optical absorption studies of colour centres in the alkali halides have led to relatively precise information on the properties of trapped electrons and positive holes in these crystals (Seitz 1954). However, for zinc and cadmium sulphides and other luminescent solids, data on trapped carriers are usually obtained from phosphorescence and thermoluminescence experiments (Garlick 1949). More recently studies have been made on single crystal specimens of zinc and cadmium sulphides of the changes in electrical conductivity during phosphorescence and thermoluminescence (Broser and Broser-Warminsky 1955). These provide an alternative means of investigating the trapping processes. Changes in absorption spectra of phosphors on excitation offer further means of obtaining information on energy storage in phosphors. Lenard (1928) found such absorption in the 0·43 to 0·56 μ region for ZnS–Cu phosphors. Hoch (1940) extended measurements to 1 μ finding in ZnS–Cu and unactivated ZnS a broad absorption band over the whole range. For ZnS–Ag phosphors excitation gave absorption bands with several maxima. Unfortunately no absolute values of absorption are given in his paper. Urbach *et al.* (1946) found that during excitation of the infra-red stimulable phosphor SrS–Eu–Sm new absorption bands appeared. The peaks at 0·6 and 1·0 μ are characteristic of the samarium impurity. It seems probable that the absorption is associated with trapped electrons though, as discussed below, the optical transitions involved do not necessarily occur in the trapping states.

This paper reports measurements on absorption bands produced in a single crystal of CdS by excitation.

---

† On study leave from the Hebrew University, Jerusalem.

§ 2. Experimental Apparatus and Technique

2.1. *The Monochromator System*

Monochromatic light from a small Hilger constant deviation quartz prism monochromator (figure 1) was split into two convergent beams by a lens and mirror system. One beam (to be referred to hereafter as the *measuring beam*) came to a focus at the phosphor crystal and the other at an aperture of similar size to that of the crystal. The emergent divergent beams from crystal and aperture were converged by separate lenses on to the same area of a suitable radiation detector passing through a *chopper disc* just in front of the detector. The incident beams were chopped so that their resultant modulation (800 c/s) differed in phase from each other by 180°. By means of an attenuator sector in the beam splitter the relative intensities of the beams could be altered. When the beams were of equal intensity no signal was recorded. The chopper disc also provided a reference signal 90° out of phase with that of either beam. After amplification of the detector signal by an 800 c/s tuned amplifier it was compared with this reference signal in a phase sensitive rectifier network and the final output to a galvanometer or to the Y plates of a cathode-ray tube then gave the *difference in intensity* of the two light beams. The detector could be one of three devices: a photomultiplier with antimony–caesium surface for the visible, a caesium surfaced photocell for the long wavelength visible and near infra-red regions and a lead sulphide photoconductive cell for the infra-red region up to $3\,\mu$. The monochromator slit width was about $0{\cdot}015\,\mu$ at $0{\cdot}5\,\mu$ and $0{\cdot}05\,\mu$ at $1{\cdot}0\,\mu$ wavelength.

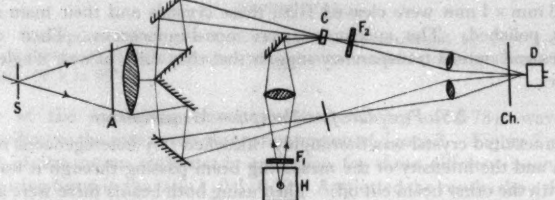


Figure 1. Apparatus for measurement of optical absorption in CdS crystals. S, exit slit of monochromator; A, beam attenuator; H, mercury arc light source; F_1, filter giving 5461 Å radiation; C, CdS crystal; F_2 filter removing 5461 Å radiation from measuring beam; Ch, chopper disc; D, radiation detector.

2.2. *The Crystal Cryostat*

The crystal was mounted in a metal Dewar system with heating coil and thermocouple and with optical access to both sides of the crystal for the measuring beam and also for the exciting radiation. Suitable filters were provided to prevent the latter from passing into the detector system with the measuring beam. Small temperature differences occurred across the crystal during warming experiments but these were not serious enough to upset experiments. They were minimized by filling the cryostat with dry air after evacuation to remove unwanted vapours.

2.3. *Light Sources*

The light source for the monochromator was a 50-watt tungsten filament lamp with variable intensity control. For excitation of the phosphor crystal a 125-watt compact source, mercury-in-quartz lamp was used with filters to select the 5461 Å line of mercury since this gives the optimum increase in absorption. An auxiliary tungsten lamp with Wratten No. 87 filter was used to provide intense heterogeneous infra-red irradiation and consequent '*bleaching*' of the excited absorption bands of the crystals before experiments.

2.4. *Crystal Specimens*

Cadmium sulphide crystals made from precipitated sulphide by heating in sealed evacuated silica tubes at 1050°C were used. No impurity activators were added but traces of chlorine were present. The crystals were rather redder in colour than those of near stoichiometric constitution. Specimens of size 4 mm × 3 mm × 1 mm were cleaved from these crystals and their main surfaces optically polished. The specimens were non-luminescent. Their cleavage properties and optical transparency suggest that they were of near single crystal character.

2.5. *Procedure for Absorption Measurements*

The unexcited crystal was thoroughly 'bleached' by heterogeneous infra-red radiation and the intensity of the measuring beam passing through it was determined with the other beam cut off. Then using both beams these were adjusted to be of equal intensity. The crystal was then excited with 5461 Å radiation. The resulting absorption decreases the intensity of the measuring beam at the detector, the resulting signal output giving the difference in intensity of the two beams. Comparison with the original intensity of the measuring beam gives the percentage change in crystal transmission and hence the absorption at the selected wavelength of the measuring beam.

§ 3. EXPERIMENTAL RESULTS

3.1. *The Absorption Spectra due to Excitation*

The absorption spectra during continuous excitation of a CdS crystal at 90°K and 290°K are shown in curves *a* and *b* of figure 2. The maximum absorption is seen to be about 10% at 90°K and is stable when excitation is removed and if the measuring beam is cut off. The latter 'bleaches' the absorption band and so, except in deliberate bleaching experiments, is kept at as low an intensity as possible while an absorption measurement is being made. The curves *a* and *b* are for a crystal which has been carefully annealed after preparation. If, however, the specimen is reheated to 600°K and quickly cooled the absorption spectrum due to excitation is enhanced and occurs right up to the absorption edge of the specimens as shown in curve *c* of figure 2. The spread of experimental points near the absorption edge is more than that due to experimental error and suggests a fine structure. The latter is analogous to that found in the 'edge' emission spectrum of luminescence which lies in this region (Klick 1953, Kroger 1955). Further studies of the effect of heat treatment on the absorption are in progress.

3.2. *Bleaching Effects of the Measuring Beam*

Since the measuring beam removes the absorption due to excitation, experiments were made to determine the decrease in absorption with time and with

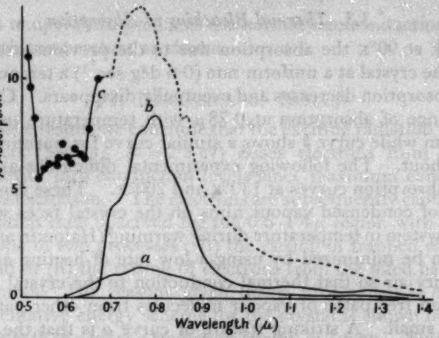


Figure 2. Absorption bands in a CdS crystal due to excitation by 5461 Å radiation. *a*, at 290°K; *b*, at 90°K; *c*, absorption due to excitation of crystal cooled quickly from 600°K to 90°K.

intensity of the measuring beam, the latter being of 0·78 μ wavelength, corresponding to the maximum absorption region of figure 2. Figure 3 gives the variation of absorption with steady excitation for three different intensities of the measuring beam in the ratio 1 : 0·5 : 0·13. As plotted the results show that there is a linear relation between the inverse of absorption $1/A$ and the inverse of the excitation intensity $1/J$ of the form

$$1/A \propto C/J + 1 \qquad \ldots\ldots(1)$$

where C is approximately proportional to the intensity of the measuring beam.

With excitation removed the absorption decays at a rate determined by the measuring beam intensity. Figure 4 gives the decay curve for different values of this intensity. The curves can be analysed into components of constant slope. The initial slope is proportional to the measuring beam intensity but later components appear to be approximately independent of the latter.

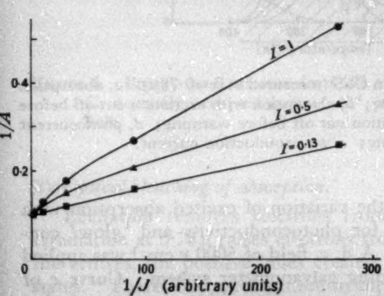


Figure 3. Variation of absorption A of CdS at 0·78 μ and 90°K with excitation intensity J showing inverse relation for three different intensities I of measuring beam.

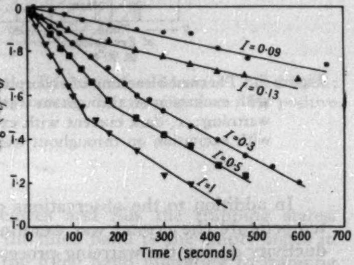


Figure 4. Optical bleaching decay curves for absorption of CdS at 0·78 μ and 90°K for different intensities of measuring beam (I, $\lambda = 0{\cdot}78\,\mu$); A_0 is absorption at $t=0$, A_t that at time t.

3.3. *Thermal Bleaching of Absorption*

In the dark at 90°K the absorption due to the previous excitation is stable. On warming the crystal at a uniform rate (0·6 deg sec^{-1}) a temperature is reached at which the absorption decreases and eventually disappears. Curve *a* of figure 5 shows the change of absorption at 0·78 μ with temperature using a very weak measuring beam while curve *b* shows a similar curve for warming with excitation left on throughout. The following experimental difficulties are found. Peaks appear in the absorption curves at 130°K and 200°K. These can be traced to the optical effects of condensed vapour films on the crystal faces which lag behind the rest of the system in temperature during warming (Halperin and Garlick 1954). The effects can be minimized by using a low rate of heating and by filling the cryostat with dry air so that thermal conduction to the crystal is improved and so that the mean free paths of vapour molecules from other parts of the system are made very small. A striking feature of curve *a* is that the absorption rises abruptly when excitation is cut off at 90°K before warming. This effect is a real increase in absorption and is not due to removal of exciting radiation which might be leaking on to the detector through the optical filter which cuts such radiation. However, no adequate explanation of the effect has yet been obtained.

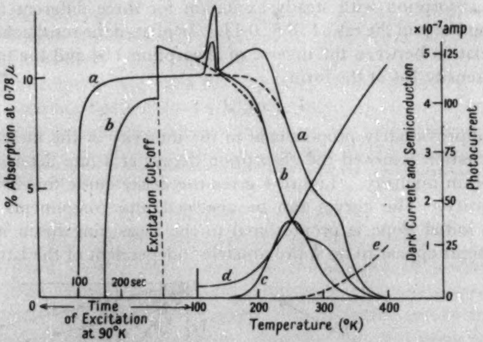


Figure 5. Thermal bleaching of absorption in CdS (measured at $\lambda = 0{\cdot}78\,\mu$). *a*, absorption with excitation on throughout warming; *b*, absorption with excitation cut off before warming; *c*, dark current with excitation cut off before warming; *d*, photocurrent with excitation on throughout warming; *e*, semiconduction current.

In addition to the observations on the variation of excited absorption with temperature the crystal was examined for photoconductivity and 'glow' conductivity during the warming process. A d. c. field of 5000 v cm^{-1} was applied and conductivity measured by a recording galvanometer system. Curve *c* of figure 5 gives the variation of photoconduction current with temperature during warming and curve *d* the current measured when the crystal was warmed in the dark as for curve *a*. The curves were not dependent on electric field direction. It is clear that the fall in absorption with temperature is closely associated with the release of stored charge carriers. Following the peak due to release of carriers a further rise in temperature produces the normal semiconduction current which rises rapidly with temperature as shown.

§ 4. Discussion of Results

From the above results we conclude that the exciting radiation giving optimum absorption effects (5400 Å; $h\nu = 2{\cdot}3$ ev) raises electrons from specific centres in the crystal lattice into conduction levels since maximum conductivity is also observed for this excitation. The electrons are subsequently captured in trapping states. Absorption then observed for the 0·6–1·4 μ region with a peak at 0·78 μ can be due to one of two processes: (i) the raising of trapped electrons into the conduction band or (ii) the raising of electrons from the filled band of the lattice into empty centres. For reasons given below process (ii) is considered more probable. A tentative energy scheme for the crystal is shown in figure 6. The absorption peak at 0·78 μ gives a transition energy of 1·6 ev. Since in the thermal bleaching of absorption shown in figure 5 there is an accompanying dark current it is possible from the exponential rise with temperature of the initial portion of the current curve *c* of figure 5 to obtain the activation energy which is about 0·18 ev. Unless a very large Franck–Condon shift is allowed, which is unlikely (Mott and Gurney 1940), the 1·6 ev transition would not appear to be the same as that involved in thermal activation. The following processes are therefore suggested:

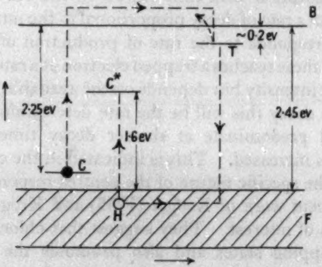


Figure 6. Energy level scheme for CdS crystal. B, conduction band; F, filled band; C, normally occupied centre; C*, empty centre; T, trapping state; H, positive hole.

(i) *Optical bleaching of absorption.*

Excitation removes electrons from centres and fills the trapping states. Irradiation at 0·78 μ raises electrons from the filled band to the empty states in the centres, the positive holes created in the filled band migrating to trapping states. There they are captured in the coulomb fields of trapped electrons and combine with the electrons thus returning the crystal to its unexcited state.

(ii) *Thermal bleaching of absorption.*

Excitation empties centres and fills traps. Thermal activation of the trapped electrons into the conduction band results in their recombination with centres thus returning the crystal to its unexcited state. In transition through the conduction band the thermally activated electrons will give rise to the current observed during thermal bleaching. The same trapping states will also affect the photoconductivity of the specimen. The latter will be small at temperatures below those for thermal bleaching and will increase rapidly as the temperature is raised to values above those for bleaching. It is of interest to note that experimentally the photocurrent rises exponentially with temperature and gives an activation energy of about 0·12 ev. The normal semiconduction of the unexcited crystal shown in curve *c* at temperatures above 300°K gives a single activation energy of 0·22 ev.

Assuming the above hypothesis it is possible to explain the relations between intensities of absorption A and of exciting J and bleaching radiation I. If there are N centres of which n are excited then there are n trapped electrons (neglecting conduction electrons in transit). Centres are emptied at a rate $\alpha J(N-n)$ by exciting radiation and replenished with electrons from the filled band excited by bleaching radiation at a rate βIn, where α and β are constants. In equilibrium

$$\alpha J(N-n)=\beta In. \qquad \ldots\ldots(2)$$

Since the observed absorption is proportional to n then by substitution in equation (2) and rearrangement

$$k/A=(k'I/J)+1 \qquad \ldots\ldots(3)$$

which agrees with equation (1) obtained from experiment.

The decay of absorption with time shown in the curves of figure 4 has an initial component with a rate of decay proportional to the intensity of the bleaching radiation. This corresponds to the rate of production of positive holes in the filled band. Each of these reaches a trapped electron at a rate which is independent of the bleaching light intensity but depends on the availablity of trapped electrons. At later stages of the decay this will be the rate determining process as shown in experiment. It will predominate at shorter decay times as the intensity of bleaching radiation is increased. This is indicated in the curves of figure 4.

With respect to the specific nature of the centres responsible for the observed absorption, some recent work of Kröger (1955) and Kröger, Vink and van den Boomgaard (1954) is of interest. They suggest that chlorine inclusion provides shallow electron trapping states and also promotes the generation of cation vacancies, which with a single positive charge function as centres with ground states 1·5 ev below the conduction band (thermal activation energy). In our results the excitation characteristics give this state at 0·2–0·3 ev above the filled band. When it has lost an electron by excitation the cation vacancy level will move to a lower energy value but to explain the above results it should move to higher energies. This would occur if the centres were monovalent cadmium ions compensated by the chlorine present. However, more precise evidence is required to distinguish the type of centre involved.

The effect observed at the absorption edge in crystals rapidly cooled after heating to a temperature of 600°K must be associated with lattice strains. It is significant that the so-called 'edge emission' of the crystal occurs in this region (Kröger 1939). Further experiments are indicated to determine possible relations between the edge emission efficiency and the thermal treatment of the crystals and also between the latter and the absorption produced in this region by excitation of such crystals.

REFERENCES

BROSER, I., and BROSER-WARMINSKY, R., 1955, *Luminescence*, Suppl. 4 to *Brit. J. Appl. Phys.*, p. S 90.

GARLICK, G. F. J., 1949, *Luminescent Materials* (Oxford : University Press).

HALPERIN, A., and GARLICK, G. F. J., 1954, *Phys. Rev.*, **95**, 1098.

HOCH, A., 1940, *Ann. Phys. Lpz.*, **38**, 485.

KLICK, C. C., 1953, *Phys. Rev.*, **89**, 274.

KRÖGER, F. A., 1939, *Physica*, 6, 779; 1955, *Luminescence*, Suppl. 4 to *Brit. J. Appl. Phys.*, p. S 58.

KRÖGER, F. A., VINK, H. J., and VAN DEN BOOMGAARD, J., 1954, *Z. phys. Chem.*, **203**, 1.

LENARD, P., 1928, *Handb. Exp. Phys.*, **23**, 690.

MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford : University Press).

SEITZ, F., 1954, *Rev. Mod. Phys.*, **26**, 7.

URBACH, F., PEARLMAN, D., and HEMMENDINGER, H., 1946, *J. Opt. Soc. Amer.*, **36**, 373.

The A.C. Admittance of Temperature-Dependent Circuit Elements

By R. E. BURGESS†
Department of Scientific and Industrial Research, Radio Research Station, Slough

Communicated by R. L. Smith-Rose; *MS. received 7th October* 1954 *and in revised form* 17*th March* 1955

Abstract. The relation between the isothermal and steady-state current–voltage characteristics of a device in which the current depends on both the voltage and the temperature is deduced quite generally without assuming electrical linearity or Newtonian cooling. The phenomena of voltage turnover in a device with a positive temperature coefficient of d.c. conductance, and of current turnover in a device with a negative coefficient are deduced.

The a.c. admittance for a small a.c. voltage superimposed on the d.c. voltage determining the operating point is derived in terms of the thermal admittance of the device, i.e. the complex ratio of the alternating components of power and temperature. For the special case in which the thermal admittance consists of a shunt combination of a fixed thermal capacity and thermal conductance the electrical admittance has been evaluated in detail. Three-element equivalent circuits (of two resistors and a capacitor or inductor) can represent the electrical admittance of such a device and its locus is semicircular; the reactive element represents the effect of thermal inertia and its nature is determined by the sign of the temperature coefficient of d.c. conductance.

Where heat loss from the device is by a path which is not 'thermally short' at the frequency under consideration the thermal admittance no longer has a simple form but has to be determined from the diffusion equation. This is done for the case of one-dimensional heat conduction (e.g. along the supporting wires of a thermistor or a lamp) or for radial heat flow (e.g. from the contact in a point-contact rectifier). The general condition for the admittance locus still to be circular, but with the centre not on the conductance axis, is formulated.

§ 1. Formulation of the Problem

Most solid-state circuit elements have a d.c. resistance which is a function of the ambient temperature; examples of these are barretters, bolometers, filament lamps, thermistors, varistors, surge protectors, crystal rectifiers and mixers. The first implication of this temperature dependence is that there are two distinct d.c. current–voltage characteristics—the isothermal and the steady state. The isothermal characteristic relates the current and voltage in conditions in which the electrical power dissipated in the device is prevented from producing any rise of temperature above the ambient temperature; this might be achieved for instance by the use of a pulsed voltage or by a sufficient increase of the cooling arrangements. For the steady-state characteristic the current is determined when it has reached its final equilibrium value after the application of voltage, and correspondingly the temperature rise produced by the electric power has attained its steady value.

---

† Now at Department of Physics, University of British Columbia, Vancouver.

At this stage it is necessary to distinguish between linear and non-linear devices; these terms will be applied to the isothermal characteristic so that if a device is non-linear when its temperature is held constant, the non-linearity is not due to heating effects but to other phenomena such as field effects or barrier layers. Thus the thermistor is a linear device in the present sense since although the steady-state characteristic is non-linear it nevertheless obeys Ohm's law in the isothermal condition and a well-defined resistance can be assigned at each temperature.

The present analysis is concerned with the apparent a.c. impedance of a temperature-dependent circuit element (linear or non-linear) when a small a.c. voltage is applied in addition to the d.c. bias which determines the operating point. As expected, the presence of thermal inertia i.e. the lag of temperature rise behind the variation of applied power will give rise to reactive impedances which may be capacitive or inductive. Furthermore the a.c. resistance will in general be different from the isothermal or the steady state differential resistance. In some circumstances the a.c. resistance can become zero or negative and the phenomenon may occur of a plateau in the characteristic or of 'turnover' and, in the right conditions of the external circuit, of oscillations. Some earlier work on linear systems only has been carried out by Smith (1950) and Stuckes (1953).

§2. GENERAL EQUATIONS

Let V and I be the instantaneous voltage and current applied to the element whose instantaneous temperature is T. An essential feature of the present analysis is that the (current, voltage) relation is determined by a single temperature T and does not depend on temperature gradients or on the existence of diverse temperatures in different regions of the device. This assumption is readily justified when the temperature of the essential current-carrying regions are substantially uniform as in a lamp or a thermistor but it will require closer examination when very localized heating occurs as for instance in the point-contact rectifier. In this latter case the characteristic temperature which defines the electric properties of the device would be taken as that at the contact or that of the region in the bulk which is the source of minority carriers.

If the rate of loss of heat is determined only by the instantaneous excess temperature of the element, the relation for balance of power is

$$C\frac{d\theta}{dt}+f(\theta)=P\equiv IV \qquad \ldots\ldots(1)$$

where $\theta=T-T_a$ is the excess of the temperature T of the element above the ambient temperature T_a. $f(\theta)$ is the rate at which heat is lost by conduction, convection or radiation. C is the heat capacity of the device at T and will be assumed constant. Only if Newton's law of cooling is valid would $f(\theta)$ be proportional to θ.

The current, voltage and excess temperature will all be assumed to have steady components with relatively small superimposed a.c. components:

$$\left.\begin{aligned} I&=I_0+I_1\exp j\omega t\\ V&=V_0+V_1\exp j\omega t\\ T-T_a=\theta&=\theta_0+\theta_1\exp j\omega t \end{aligned}\right\} \qquad \ldots\ldots(2)$$

Clearly I_0 and V_0 must define a point on the steady-state characteristic for ambient temperature T_a and must also be on the isothermal for temperature $T_a+\theta_0$. In general $I_1 V_1$ and θ_1 are complex quantities. The power P is then

$$P \equiv P_0 + P_1 \exp j\omega t = IV = I_0 V_0 + (I_0 V_1 + I_1 V_0) \exp j\omega t \qquad \ldots\ldots(3)$$

to the present degree of approximation. Substitution in equation (1) gives

$$f(\theta_0) = I_0 V_0$$

$$\theta_1 y \equiv \theta_1 [j\omega C + f'(\theta_0)] = P_1 = I_0 V_1 + I_1 V_0. \qquad \ldots\ldots(4)$$

Here $f'(\theta_0)$ is the thermal conductance of the cooling mechanism, and P_1/θ_1 is the total thermal admittance y.

Finally the alternating component of current is determined by both V_1 and θ_1 through the isothermal a.c. conductance g and the current temperature coefficient h:

$$I_1 = \left(\frac{\partial I}{\partial V}\right)_T V_1 + \left(\frac{\partial I}{\partial T}\right)_V \theta_1 \equiv g V_1 + h\theta_1. \qquad \ldots\ldots(5)$$

In general g and h are functions of temperature. For a linear device, however, g is independent of voltage and thus $g = I_0/V_0$ for this case. The parameter h distinguishes by its sign the two essential types (figure 1) of steady-state characteristics: (i) that which has everywhere a slope greater than the isothermal at the same point (if $h>0$) or (ii) that which everywhere has a slope less than the isothermal ($h<0$).

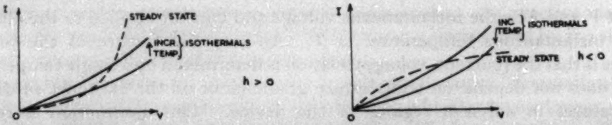


Figure 1. Isothermal and steady-state current–voltage characteristics of temperature-dependent circuit elements. In lineal devices the isothermals are straight lines.

The quantity

$$\tau \equiv C/f'(\theta_0) \qquad \ldots\ldots(6)$$

is the effective thermal time constant of the device which in general will be a function of the excess temperature θ_0, unless Newton's law of cooling is valid.

The complex ratio of the alternating components of temperature and voltage is

$$\frac{\theta_1}{V_1} = \frac{I_0 + gV_0}{y - hV_0} \qquad \ldots\ldots(7)$$

and the a.c. admittance is given by

$$Y(\omega) = \frac{I_1}{V_1} = \frac{gy + hI_0}{y - hV_0}. \qquad \ldots\ldots(8)$$

At zero frequency the admittance is a pure conductance and the steady-state characteristic is described by the equation

$$Y(0) = G(0) = \frac{gy_0 + hI_0}{y_0 - hV_0} \equiv G_0 \qquad \ldots\ldots(9)$$

where the thermal admittance y_0 at zero frequency is a real quantity. It may readily be seen that G_0 is simply the slope of the steady-state current–voltage characteristic as is to be expected physically, for if ΔI, ΔV and $\Delta\theta$ are corresponding (infinitely slow) increments along the steady-state characteristic

$$y_0 \Delta\theta = I_0 \Delta V + V_0 \Delta I \qquad \text{and} \qquad \Delta I = g\Delta V + h\Delta\theta$$

leading to the value of $G_0 \equiv \Delta I/\Delta V$ given in the preceding equations.

At infinite frequency the admittance becomes

$$Y(\infty)=G(\infty)=g \qquad \text{......(10)}$$

which is the isothermal conductance, appropriate here since the temperature of the device cannot follow the infinitely rapid variation of electrical power.

In the case where the thermal admittance y has the form $y=j\omega C+C/\tau$ the conductance and susceptance at angular frequency ω are:

$$\left.\begin{aligned} G(\omega) &= \frac{g\omega^2C^2+(gC/\tau+hI_0)(C/\tau-hV_0)}{\omega^2C^2+(C/\tau-hV_0)^2} \\ &= g+\frac{h(I_0+gV_0)(C/\tau-hV_0)}{\omega^2C^2+(C/\tau-hV_0)^2} \end{aligned}\right\} \qquad \text{......(11)}$$

$$B(\omega)=-\frac{\omega Ch(gV_0+I_0)}{\omega^2C^2+(C/\tau-hV_0)^2}. \qquad \text{......(12)}$$

The corresponding formulae for the resistive and reactive components of the impedance are:

$$R(\omega)=\frac{g\omega^2C^2+(C/\tau-hV_0)(gC/\tau+hI_0)}{g^2\omega^2C^2+(gC/\tau+hI_0)^2} \qquad \text{......(13)}$$

$$X(\omega)=\frac{\omega Ch(gV_0+I_0)}{g^2\omega^2C^2+(gC/\tau+hI_0)^2}. \qquad \text{......(14)}$$

The time variation of the current and voltage due to the (relatively small) a.c. voltage superimposed on the d.c. bias is illustrated in figure 2 for the case of $h>0$. At very low frequency ($\ll 1/\tau$) the temperature is able to attain its steady value at each instant in accordance with the applied power IV and accordingly the locus is a small arc (essentially a straight line in the above analysis) of the steady-state characteristic so giving a pure conductance of G_0. At a very high frequency ($\gg 1/\tau$) however, the power is varying so rapidly that the temperature remains constant at the value determined by the average applied power; in this case the locus is an arc of the isothermal characteristic and corresponds to a pure conductance equal to g. At an intermediate frequency the temperature somewhat sluggishly follows the applied power variation but does so with a lag corresponding to the phase angle of the thermal admittance $y(\equiv P_1/\theta_1)$. The locus is therefore an ellipse and simple considerations readily show that it is traced out in an

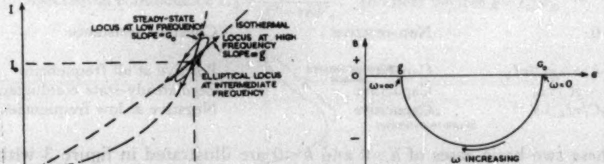

I
STEADY-STATE LOCUS AT LOW FREQUENCY SLOPE = G₀
ISOTHERMAL
LOCUS AT HIGH FREQUENCY SLOPE = g
I₀
ELLIPTICAL LOCUS AT INTERMEDIATE FREQUENCY
V₀
V
(a)
B
+
0
−
g
G₀
G
ω = ∞
ω = 0
ω INCREASING
(b)

Figure 2. (*a*) Current–voltage locus for small a.c. voltage superimposed on d.c. bias (illustrated for case $h>0$). The isothermal characteristic is for the temperature T_0 of the device so that it passes through the operating point (I_0, V_0).
(*b*) Locus of admittance for the case $h>0$.

anti-clockwise sense. The case of $h<0$ is similar except that $G_0<g$ and the ellipse is traced out in a clockwise sense.

The phenomenon of 'turnover' in the steady-state characteristic is seen to correspond as follows: if $h>0$, $y_0=hV_0$ (max) at the turnover voltage V_0 (max), as in a thermistor; if $h<0$, $gy_0=-hI_0$ (max) at the turnover current I_0 (max), as in a barretter.

§ 3. PROPERTIES OF THE ADMITTANCE AND IMPEDANCE FUNCTIONS

3.1. *General*

It is immediately seen from the form of the general admittance equation (8) that if g, c, h and τ can be regarded as independent of frequency the admittance and impedance have a semicircular locus in the complex plane with the zero- and infinite-frequency values (both real) at the ends of the diameter. Correspondingly this suggests a number of three-element equivalent circuits containing two resistors and one reactor, all elements being independent of frequency. The nature of the reactance and resistance is determined by the following simple rules given in the table.

Character of the A.C. Admittance

Limits of h	Reactive Character	Resistive Character
$h>C/\tau V_0$	Inductive	Negative at low frequencies
$h=C/\tau V_0$	Inductive	Zero steady-state resistance
$C/\tau V_0>h>0$	Inductive	Positive at all frequencies
$h=0$	Non-reactive	Constant resistance
$0>h>-gC/\tau I_0$	Capacitive	Positive at all frequencies
$h=-gC/\tau I_0$	Capacitive	Zero steady-state conductance
$-gC/\tau I_0>h$	Capacitive	Negative at low frequencies

These two basic cases of $h>0$ and $h<0$ are illustrated in figure 3 with two forms of equivalent circuit for each case.

It is seen that if h lies outside the limits $+C/\tau V_0$ and $-gC/\tau I_0$ the device will display negative resistance up to a critical frequency given by

$$\omega_0^2=\left(\frac{hV_0}{C}-\frac{1}{\tau}\right)\left(\frac{hI_0}{gC}+\frac{1}{\tau}\right). \qquad \text{......(15)}$$

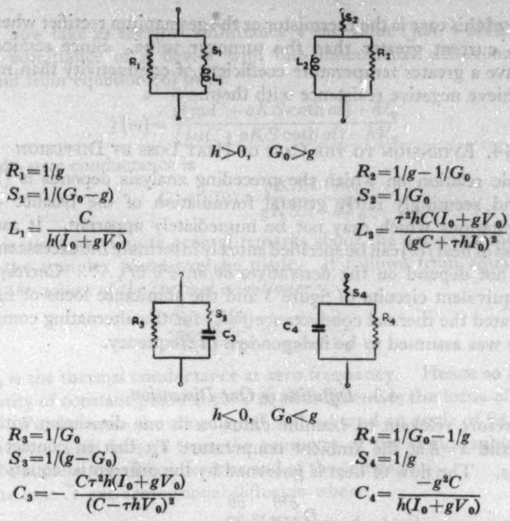
At this frequency the resistance and conductance are zero and the admittance is a pure susceptance given by

$$Y(\omega_0)=jB(\omega_0)=j(-G_0g)^{1/2} \qquad \text{......(16)}$$

as may be readily seen from the geometry of the admittance locus.

In these conditions G_0 is negative; for $h>0$ the negative root must be taken and if $h<0$, the positive root.

A device displaying negative resistance in the steady state characteristic may be classified as either current-controlled or voltage-controlled (figure 4) and these correspond to $h>C/\tau V_0$ or $h<-gC/\tau I_0$ respectively. The types of resonant circuit with which they must be used to provide oscillation are respectively series and parallel. However it must be remembered that the thermal inertia of



$$h>0, \quad G_0>g$$

$$R_1=1/g \qquad R_2=1/g-1/G_0$$

$$S_1=1/(G_0-g) \qquad S_2=1/G_0$$

$$L_1=\frac{C}{h(I_0+gV_0)} \qquad L_2=\frac{\tau^2hC(I_0+gV_0)}{(gC+\tau hI_0)^2}$$

$$h<0, \quad G_0<g$$

$$R_3=1/G_0 \qquad R_4=1/G_0-1/g$$

$$S_3=1/(g-G_0) \qquad S_4=1/g$$

$$C_3=-\frac{C\tau^2h(I_0+gV_0)}{(C-\tau hV_0)^2} \qquad C_4=\frac{-g^2C}{h(I_0+gV_0)}$$

Figure 3. Equivalent circuits for temperature-dependent circuit elements. Steady-state differential conductance $G_0=\dfrac{gC+\tau hI_0}{C-\tau hV_0}$. In linear devices $g=I_0/V_0$.

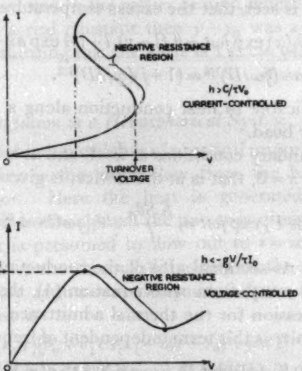


Figure 4. Negative resistance regions of the characteristics of sufficiently temperature-dependent circuit elements.

the device endows it with reactance so in principle it may only be necessary to connect in circuit a reactance of the opposite sign and of suitable magnitude to give oscillation. This is a valuable practical feature in the case of the inductive class of negative resistance since it is then only necessary to add capacitance to permit continuous oscillation over the frequency range for which $R(\omega)<0$.

An example of this case is the thermistor or the germanium rectifier when suitably biased to a current greater than the turnover value. Since semiconductors generally have a greater temperature coefficient of conductivity than metals it is easier to achieve negative resistance with them.

§4. EXTENSION TO THE CASE OF HEAT LOSS BY DIFFUSION

The basic relation on which the preceding analysis depends is (1). This plausible and seemingly fairly general formulation of the balance of power contains limitations which may not be immediately apparent. It asserts that the rate of loss of heat $f(\theta)$ can be specified entirely in terms of the excess temperature θ and does not depend on the derivatives $\partial\theta/\partial t$, $\partial^2\theta/\partial t^2, \ldots$. Correspondingly when the equivalent circuits of figure 3 and the admitance locus of figure 2 (*b*) were formulated the thermal conductance $f'(\theta_0)$ for the alternating component of temperature was assumed to be independent of frequency.

4.1. *Diffusion in One Dimension*

It is therefore relevant to examine diffusion in one dimension with $x=0$ at the device, and $x=l$ at the ambient temperature T_a, that is, contact with the surroundings. The flow of heat is governed by the differential equation

$$D\frac{\partial^2\theta}{\partial x^2}=\frac{\partial\theta}{\partial t}$$

where the thermal diffusivity $D=K/\rho s$ in which K is the thermal conductivity of the conducting medium providing the cooling path, ρ is its density and s its specific heat. It is seen that the excess temperature $\theta(x)$ at x will be of the form

$$\theta(x)=\theta_0(x)+\theta_1(x)\exp j\omega t=\theta_0(l-x)/l+[A\exp ax+B\exp(-ax)]\exp(j\omega t)$$

where $$a\equiv(j\omega/D)^{1/2}=(1+j)(\omega/2D)^{1/2}. \qquad \ldots\ldots(17)$$

This case is applicable to heat conduction along a wire (or wires) for example from a thermistor bead.

From the boundary conditions at $x=0$ and at $x=l$ it follows that the instantaneous power at $x=0$, that is at the device, is given by

$$P=P_0+P_1\exp j\omega t=\frac{KS\theta_0}{l}+\theta_1[j\omega C+aKS\coth al]\exp(j\omega t)$$

where S is the cross-sectional area of the conduction path.

It is seen, by comparison with equation (4), that aKS coth al has replaced $'(\theta_0)$ in the expression for the thermal admittance P_1/θ_1. Only if $|al|$ is small compared with unity is this term independent of frequency and in this case

$$l(\omega/D)^{1/2}\ll 1, \qquad aKS\coth al=KS/l=P_0/\theta_0$$

which is simply the d.c. thermal conductance of the cooling path. If on the other hand l or ω are sufficiently large

$$l(\omega/D)^{1/2}\gg 1, \qquad aKS\coth al=(1+j)(\omega/2D)^{1/2}KS$$

which is complex and increases as $\omega^{1/2}$. This means that the thermal time constant τ of equation (6) can no longer be regarded as a real constant in equations (11)–(14) although the general equations (7)–(10) are still valid.

If we now take as thermal admittance y the value $(j\omega C+aKS\coth al)$, the electrical admittance of a device with one-dimensional diffusional cooling is then found from equation (8) to be

$$Y(\omega)=\frac{g(j\omega C+aKS\coth al)+hI_0}{(j\omega C+aKS\coth al)-hV_0}. \qquad \ldots\ldots(18)$$

The steady-state conductance is

$$G_0=Y(0)=\frac{gKS/l+hI_0}{gKS/l-hV_0}.$$

At this point some quite general remarks about the nature of the admittance locus in the complex plane may be made. It follows from equation (8) that *whatever the nature of the thermal admittance y*:

$$\frac{G_0-Y(\omega)}{Y(\omega)-g}=\frac{y-y_0}{y_0-hV_0} \qquad \ldots\ldots(19)$$

where y_0 is the thermal conductance at zero frequency. Hence so long as $y-y_0$ is a quantity of constant phase angle ϕ at all frequencies the locus of $Y(\omega)$ will be circular. Furthermore the chord g_0, G will subtend an angle of 2ϕ at the centre of the circle. This follows from the geometry of the angles subtended by the chord of a circle at the circumference and at the centre.

In the case of one-dimensional diffusion where

$$y-y_0=j\omega C+aKS[\coth al-1/al]$$

we clearly do not have a quantity with constant phase angle and the locus of Y will not be circular.

In the simpler case considered in the earlier section where Y was $j\omega C+f'(\theta_0)$ and $f'(\theta_0)$ was regarded as a real constant then $y-y_0$ was a quantity of phase angle $\pi/2$ at all frequencies showing that the locus is a circle with g_0, G as the ends of a diameter.

4.2. *Diffusion in a Hemispherical System*

A case of heat loss by conduction which is of practical importance in the reverse biased point-contact rectifier is the radial flow from the barrier layer region through the semiconductor. Here the heat is generated in a very small hemispherical shell where it is assumed all the power is dissipated. The radius of the shell is r_0 and heat is presumed to flow out to $r=\infty$ where the excess temperature is zero. If D is the thermal diffusivity of the semiconductor, the excess temperature θ at any radius r is determined from $D\nabla^2\theta=\partial\theta/\partial t$ whose solution for the present case is

$$\theta(r)=(r_0/r)[\theta_0+\theta_1\exp\{-a(r-r_0)+j\omega t\}] \qquad \ldots\ldots(20)$$

where θ_0 is the steady component of $\theta(r_0)$ and $\theta_1\exp j\omega t$ is the alternating component just as in the preceding analysis; a is still $(j\omega/D)^{1/2}$. The system is shown in figure 5 (a).

The balance of power equation is

$$P=P_0+P_1\exp j\omega t=-2\pi r_0^2K\left(\frac{\partial\theta}{\partial r}\right)_{r_0}=2\pi r_0K[\theta_0+\theta_1(1+ar_0)\exp j\omega t].$$

At low frequencies such that $|ar_0|\ll 1$ the a.c. thermal admittance is real and equal to the d.c. thermal conductance $2\pi r_0K$. At high frequencies it becomes complex

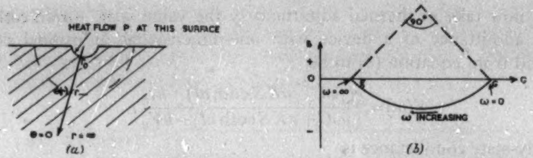


Figure 5. (*a*) Model for hemispherical diffusion of heat at a semiconductor contact.
(*b*) Locus of admittance in the Y-plane for the system considered in figure 5 (*a*).

and eventually varies as $\omega^{1/2}$. In general the thermal admittance is

$$y = P_1/\theta_1 = 2\pi r_0 K[1+ar_0] = I_0 V_0[1+ar_0]/\theta_0. \qquad \ldots\ldots(21)$$

The electrical admittance is given by

$$\left.\begin{aligned} Y(\omega) &= \frac{g[1+(1+j)\beta]+M}{1+(1+j)\beta-N} \\ \text{where} \quad M &\equiv h\theta_0/V_0,\ N = h\theta_0/I_0 \text{ and } \beta \equiv r_0(\omega/2D)^{1/2}. \end{aligned}\right\} \qquad \ldots\ldots(22)$$

The steady state conductance is $Y(0) = G_0 = (g+M)/(1-N)$ and it is therefore found that $[G_0 - Y(\omega)]/[Y(\omega)-g] = [(1+j)\beta]/[1-N]$ which shows that the locus of Y in the admittance plane is the arc of a circle, meeting the conductance axis at g and G_0 and subtending a right angle at the centre. This is to be expected from the general considerations of the preceding section since $y - y_0$ has a constant phase angle of $\pi/4$ in the present case (figure 5 (*b*)).

ACKNOWLEDGMENTS

The work described above was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

SMITH, O. J. M. 1950, *Rev. Sci. Instrum.*, **21**, 351.
STUCKES, A. D., 1953, *Proc. Phys. Soc.* B, **66**, 570.

Electroluminescence in Single Crystals of Zinc Sulphide

By G. F. ALFREY and J. B. TAYLOR
Electron Physics Department, University of Birmingham

Communicated by J. Sayers; MS. received 4th April 1955

Abstract. An apparatus for the measurement of electroluminesence in single crystals of zinc sulphide under a range of conditions is described. The results of the measurements are compared with existing theories, which are shown to be unsatisfactory. Modified forms of these theories are developed which are in much better agreement with experiment.

---

§ 1. Introduction

Since the discovery of electroluminescence by Destriau (1947) several further studies of this phenomenon have been reported and certain applications developed, in particular the so called 'electroluminescent lamp', made by dispersing a powdered phosphor in a transparent dielectric between parallel electrodes. Several investigations of the properties of these lamps have been made and in some cases certain conclusions drawn regarding the physical processes involved (Burns 1953).

However, this type of investigation is not the best starting point in the study of electroluminescence, because the dispersal of the phosphor in a dielectric adds greatly to the difficulties of interpretation. Attempts to solve this problem have been made by Roberts (1952), who estimated the field in the individual phosphor grains, and by Zalm, Diemer and Klasens (1954), who used a single layer of powder grains *in vacuo*. More satisfactory is the approach of Waymouth and Bitter (1954) who studied single grains of phosphor, but this method is limited in its scope by the very small size of the grains.

Probably the most satisfactory form of the phosphor for study is a single crystal, as used by Piper and Williams (1952), who obtained crystals about 5 mm long and 3 mm^2 in cross section. Unfortunately, single crystals are difficult to grow and the amount and nature of the activators cannot be controlled, so that crystals grown by different experimenters will differ in constitution. The results which follow refer to crystals of ZnS grown by the Frierichs method. Although these were all from the same batch, there was nevertheless a considerable deviation between the results on different crystals and between different contacts on the same crystal. We have therefore confined our attention to a discussion of those features which were present in all cases.

With electroluminescent lamps of the type described earlier the luminescence is continuous only if alternating fields are used, a static field producing only a transient burst of light when the field is switched on or off. With single crystals a continuous emission is also possible with static fields. The measurements described here were made with sinusoidal fields, a procedure which has the advantage of rendering the effect of the electrodes and the type of contact relatively unimportant.

§ 2. APPARATUS

The apparatus enabled measurements to be made over a wide temperature range. A factor in its design is the need for electrical insulation, since discharge between the electrodes would be a source of ultra-violet radiation, which in turn would produce spurious luminescence. An insulating liquid could be used but no single liquid is satisfactory over a sufficient range of temperature. Instead a high vacuum is used.

Figure 1 shows the apparatus. The vacuum enclosure is a cylindrical box with a window in one flat face through which the luminescence is observed. The crystal, in its holder, is on the opposite face. So that its temperature may be varied, the holder is carried on the inner of a pair of concentric german silver cylinders, whose low thermal conductivity insulates the holder from the rest of the apparatus. The crystal holder, which is of copper, forms a radiation shield and in addition to carrying a thermocouple also serves as one electrode. The crystal is pressed against the holder by a crimped tungsten wire which is well insulated and forms the high voltage lead.

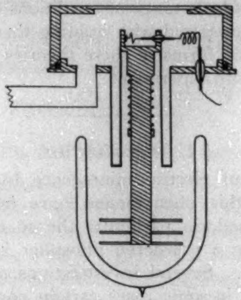


Figure 1. The apparatus.

The temperature of the crystal is controlled by a copper rod which protrudes from the back of the holder. This rod carries a heating coil wound on it near the crystal holder, and its free end can be immersed in liquid air. By this means a temperature range of 180° to -160°c can be covered.

The voltage used is variable up to 1000 volts peak to peak and from 25 c/s to 2 kc/s.

Various types of electrode were used. In the simplest, the tungsten wire formed a point contact; in another, metal electrodes were evaporated on to the crystal, and in the third case the electrodes did not make contact with the crystal, but formed a small parallel plate condenser in which the crystal was suspended.

The electroluminescence was detected with an E.M.I. type 6260 photomultiplier, using a galvanometer to measure the photocurrent and an oscilloscope to examine the wave form.

§ 3. RESULTS

As has already been stated, the use of alternating fields reduces the dependence of the results on the nature of the electrodes; in fact results of substantially the same form were obtained with all types of electrode described above. Hence the electroluminescence may be considered to arise from a process taking place within the crystal and not dependent on electrons entering from the electrodes.

It has been reported that the electroluminescence of some phosphors shows a change in wavelength as the frequency is changed (Waymouth 1953). To look for a similar effect in the present experiments, observations were made through suitable filters, which isolated the blue and green emission bands to be expected in zinc sulphide. These measurements revealed no essential difference between the results taken through the two filters and it may be concluded that there was no appreciable change in spectral character during the present experiments. In fact, these wavelength changes probably occur only in complex phosphors containing several activators.

The remaining observations can be grouped under three headings.

(i) *The dependence of electroluminescence on field strength at fixed temperature and frequency.*

It is found that the photomultiplier output B corresponding to an applied voltage V can be represented by an expression of the form

$$B = a \exp(-bV^{-1/2}) \qquad \ldots\ldots(1)$$

where a and b are constants. A graph of $\log B$ against $V^{-1/2}$ is linear over a $10^3:1$ range of light intensity (figure 2). The gradient varies slightly with temperature and frequency. Diemer and Zalm (private communication) obtain a similar result for a specimen of ZnS in powder form over a range of $10^5:1$. Destriau (1947) and Piper and Williams (1954) present their data differently, but if their points are replotted on the above basis they are also found to satisfy a relation of the form (1).

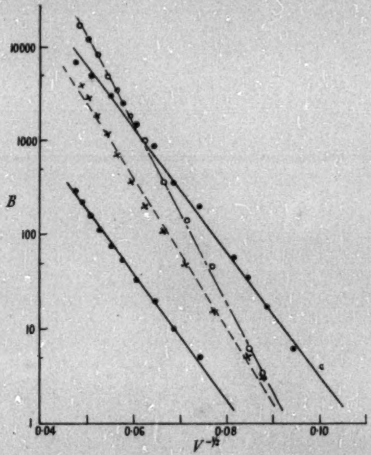


Figure 2. Variation of electroluminescence with applied voltage.

In some cases the results were more accurately represented by the sum of two expressions of the form (1); this indicates that appreciable emission is occurring at both electrodes and that owing to their asymmetry the two electrodes have different values for the constants *a* and *b*. More usually the asymmetry is so marked that the emission from one electrode is negligible. This view is supported by observations of the light emitted in alternate half-cycles, as indicated on the oscilloscope.

(ii) *The effect of temperature on electroluminescence.*

Considerable insight into the mechanism of electroluminescence can be gained from studies of the temperature variation. The general form of this is shown in figure 3, and by far the most striking feature is a large and sudden rise in the light output in the neighbourhood of 0°c. Below and above this region the variation with temperature is relatively slow.

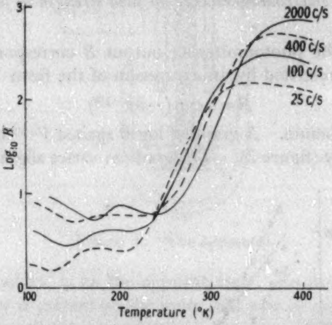


Figure 3. Effect of temperature on electroluminescence.

An increase in voltage merely increases the light output at all temperatures leaving the shape of the curve unchanged.

Repetition of these experiments at different frequencies produces a characteristic effect, an increase in frequency displacing the position of the 'step' in light output towards higher temperatures, to an extent roughly proportional to the ratio of the new frequency to the old. This effect has been observed over a range from 25 c/s to 2 kc/s as indicated in figure 3.

Published information on the variation of electroluminescence with temperature is mainly restricted to powdered phosphors. Some results do not agree with our findings (Gobrecht, Hahn and Gumlich 1954), but results similar to ours were reported by Waymouth, Jerome and Gungle (1952), and Halstead and Koller (1954), the latter for an evaporated layer of ZnS.

(iii) *Variation of electroluminescence with frequency* (*at constant temperature*).

Figure 3 indicates that a change in frequency could produce a variety of effects according to the temperature at which observations are made. This is confirmed by direct measurements and in general the results are of a very complex and irregular form. However at high temperatures the variation of electroluminescence with frequency is much simpler; at sufficiently high voltages the light output is proportional to the frequency, but as the voltage is reduced the effect of frequency becomes less marked until at the lowest voltages the light produced is practically independent of the frequency.

Various authors have claimed that the output of electroluminescent lamps is proportional to the frequency of the exciting voltage under all conditions, but these results are probably due to the change with frequency of the impedance of the matrix in which the powder is suspended and are not characteristic of the phosphor itself.

§ 4. THE MECHANISM OF ELECTROLUMINESCENCE

Piper and Williams (1954) have considered three possible mechanisms for electroluminescence and conclude that the only acceptable one is the impact excitation of luminescent centres by electrons accelerated in the field.

To account for the appearance of electroluminescence at the low average fields at which it sometimes occurs, it must be assumed that a high local field exists in a barrier at the surface of the crystal (the alternative hypothesis of a long mean free path has been disproved by the work of Ince (1954)). The presence of such a barrier also accounts for the localization of the luminescence near the electrodes. Barriers of this type are well known in rectifier theory; so far as is known at present the barrier in zinc sulphide may be of either the 'physical' or the 'chemical' type.

If the impact excitation mechanism is accepted it remains to decide the origin of the electrons which are accelerated by the field. As direct excitation of electrons from the filled band to the conduction band is impossible at reasonable fields, we must assume that the electrons originate from localized levels in the forbidden band. There are two ways in which electrons might be released from these traps: by field emission or by thermal excitation. The implications of these alternative mechanisms in the theory of electroluminescence are discussed in detail below.

4.1. *Field Emission*

A theory of electroluminescence based on field emission from donor levels (traps) has been developed by Piper and Williams (1954), who deduced an expression for the dependence of light output on applied voltage. This theory is in reasonable agreement with experiment if a suitable trap distribution is chosen. The agreement may be improved by making modifications suggested by Taylor (1954) which lead to the expression $B=aV^n \exp(-bV^{-1/2})$, indistinguishable from that found empirically in the present experiments and no longer critically dependent on the donor distribution.

However, the theory in this form makes no predictions regarding the variation with temperature, and suggests that the light output should be proportional to the frequency of the exciting voltage which, as we have seen, is the exception rather than the rule. Failure of this theory to account for the temperature effects suggests that thermal emission might be more important than field emission.

4.2. *Thermal Emission*

The probability of ejection of an electron from a trap of depth E at a temperature of T°K is usually assumed to be $\gamma = s \exp(-E/\boldsymbol{k}T)$. If this process is the source of electrons producing electroluminescence it will clearly be most effective when the traps are emptied in a time comparable with a half-cycle of the applied voltage, i.e. when

$$\omega \simeq 2\pi s \exp(-E/\boldsymbol{k}T). \qquad \ldots\ldots(2)$$

Thus the number of electrons released, and hence the electroluminescence, will increase rapidly with rising temperature, until the above condition is satisfied and all the traps are effective. This accounts qualitatively for the 'step' in the observed temperature curve. The shift of the curve towards higher temperatures when the frequency is increased is also explained. Clearly if ω is increased then condition (2) will only be satisfied at a higher temperature. Hence this hypothesis accounts qualitatively for the main features of the observed temperature dependence. A detailed analysis (Taylor 1954) puts this agreement on a quantitative basis and leads to the approximate result

$$B = \gamma \exp(-\gamma\pi/2\omega) \exp(-\beta V^{-1/2}) \qquad \ldots\ldots(3)$$

(which result is valid provided $\gamma V/\beta\omega < 1$).

It will be noticed that this theory predicts a variation of electroluminescence with voltage which agrees with that found experimentally. However, like the field emission theory, it fails to account for the observed dependence of electroluminescence on frequency.

The result (3) contains implicitly two parameters s and E and these can be chosen to give agreement with the results at any one frequency, but it has not been possible to obtain agreement with the results at all frequencies with a single pair of values for s and E. Compromise values do not give good agreement with experiment (figure 4). The theory predicts a rapid fall in electroluminescence at high temperatures which is not observed, indeed sometimes the light output remains practically constant in this region.

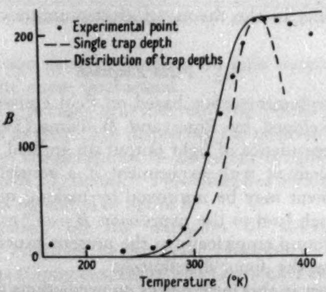


Figure 4. Comparison of thermal emission theory with experiment.

The low temperature behaviour of electroluminescence is also unexplained, for according to the theory the electroluminescence should fall to zero at very low temperatures; in fact this does sometimes occur but more usually the output becomes constant or may even increase again. However since the low temperature output is so small compared with the high temperature value this difficulty does not seem too serious in the present state of knowledge.

Agreement with experiment at high temperatures can be improved by replacing the single trap depth E by a distribution of levels.

4.3. *Thermal Emission Theory with Distribution of Trap Depths*

If, for example, we assume that there is a continuous distribution of levels below E_0, then modifying result (3), the temperature variation of electroluminescence is given by

$$B \propto \int_{E_0}^{\infty} \gamma \exp \frac{-\gamma\pi}{2\omega}\, dE = \frac{2\omega}{\pi} \boldsymbol{k}T \left[1 - \exp \frac{\gamma_0 \pi}{2\omega} \right] \qquad \ldots\ldots(4)$$

where $\gamma_0 = s \exp(-E_0/\boldsymbol{k}T)$. (The integration in equation (4) can be extended to infinity without introducing error since for large E the integrand is negligible.) This formula, which again contains two parameters s and E_0, accounts for the observed shape of the electroluminescence–temperature curve and for the shift of this curve with frequency. It still predicts that at high temperatures the light output should be proportional to frequency and, as we have seen, this simple proportionality between electroluminescence and frequency at high temperatures is only observed if the voltage is large enough.

The introduction of a distribution of trap depths merely forces agreement with experiment and is unjustified unless other evidence for it can be obtained. No such evidence has been found, for although thermoluminescence data show the presence of a trap distribution of some sort, there is no way of obtaining the form of the distribution from this, or of proving that the traps so found are those involved in electroluminescence.

Thus in spite of the initial success of the thermal emission theory it leads to an unsatisfactory situation, and we will therefore re-examine the field emission hypothesis.

§ 5. ALTERNATIVE THEORY BASED ON FIELD EMISSION

The original field emission theory considered in § 4 accounted for the dependence of electroluminescence on voltage in a reasonably satisfactory manner, but was unable to account for the temperature and frequency effects. A closer consideration of the formation of the barrier region, however, makes it possible to develop a theory free from the above limitations.

The alternating voltage applied across the crystal will in turn deplete and replenish the electron donors in the barrier region near each electrode. After a few cycles a steady state is attained in which as many electrons are withdrawn from the barrier during one half-cycle as are returned to it during the next. Removal of electrons from the barrier is the process on which attention has hitherto been concentrated, and if electroluminescence is an electron-impact phenomenon taking place in the barrier at the cathode, then the removal of electrons from this region must take place in the quarter-cycle when the cathode voltage increases from zero to its maximum. The flow can take place in this short time in spite of the high dark resistance of zinc sulphide because the electrons leaving the barrier have an energy much greater than the thermal energy of electrons in the conduction band, so that the idea of normal electrical resistance is irrelevant in describing their behaviour. During the remaining part of the voltage cycle electrons return to the barrier region from the bulk of the crystal, but under conditions which *are* controlled by the normal electrical resistance of the crystal. This return process which has previously been ignored will now be considered in detail.

If a voltage $V_0 \sin \omega t$ is applied across a crystal there will be a potential drop V_b across the barrier at the cathode. The voltage across the barrier at any time is connected with the total charge q in the barrier region by some relationship of the form $V_b = V_b(q)$.

Also if V_c is the potential drop across the bulk of the crystal

$$V_c = IR = \frac{G}{\sigma}\frac{dq}{dt} \qquad \ldots\ldots(5)$$

where G is a geometrical factor and σ is the conductivity. Then the return of charge to the barrier is governed by an equation of the form

$$V_b(q) + \frac{G}{\sigma}\frac{dq}{dt} = V_0 \sin \omega t. \qquad \ldots\ldots(6)$$

If this equation could be integrated we should know the total charge returning to the barrier. In the steady state it is only this charge which is available to produce electroluminescence, and the steady light output is governed by it. Unfortunately the integration is impossible owing to the complicated form of $V_b(q)$ but several conclusions can be drawn without the necessity for an explicit solution. Thus the form of the curve of electroluminescence against temperature can be deduced directly from equation (6). At high temperatures the conductivity will be large since zinc sulphide behaves somewhat as a semiconductor in this region, consequently the second term in the equation is negligible and $V_b(q) = V_0 \sin \omega t$, so that when $V_0 \sin \omega t$ returns to the value zero, *all* the donors have been refilled, and the electroluminescence will then have a maximum value, controlled only by the applied voltage. However, at low temperatures σ is small so the return of charge to the barrier will be seriously impeded. In fact the total charge which returns to the barrier in each cycle under these conditions cannot exceed

$$\frac{\sigma}{G}\int_0^{2\pi} (V_b(q) - V_0 \sin \omega t)\, dt$$

which is in turn less than $2\pi\sigma V_0 = q^*$ (say) since $V_b(q) \ngtr V_0$. Hence at these low temperatures the charge returned to the barrier region in each cycle is very small, and when a steady state has been attained the light output is correspondingly low, much less than the high temperature value. Since σ varies rapidly with temperature, the increase of electroluminescence towards its high temperature value is correspondingly rapid. In this way the general features of the variation of electroluminescence with temperature can be explained. The shift in the curves of electroluminescence against temperature brought about by a change in frequency can also be accounted for in terms of this theory. For this purpose the equation is rewritten in terms of $x = \omega t$ as

$$G\left(\frac{\omega}{\sigma}\right)\frac{dq}{dx} = V_b(q) - V_0 \sin x$$

the integral of which from $x = \pi/2$ to $x = 2\pi$ determines the charge which returns to the barrier. It is clear that this quantity will be of the form $q = q(x, \omega/\sigma)$ in which the effect of temperature will appear only through the ratio ω/σ.

The fact that q, and therefore B, depends on temperature only through the ratio ω/σ is exactly the condition required to give a shift in the electroluminescence–temperature curve when the frequency is changed. Following an increase in ω, B can be restored to its former value by an increase in σ which is brought about by a rise in temperature.

5.1. *Variation of Electroluminescence with Frequency*

The variation of electroluminescence with frequency only takes a simple form at high temperatures, when the electroluminescence is no longer controlled by the return of electrons to the barrier region.

In the earlier forms of the field emission theories it is assumed that all donors of depth less than some critical value E_c are ionized and that those deeper than E_c are unaffected, where E_c is determined by the maximum field. In this case the theory leads inevitably to electroluminescence proportional to frequency (Piper and Williams 1954). More detailed examination of the process of field ionization shows that there is a small probability of ionizing donors which lie below E_c. As this probability is small only a fraction of these deeper donors are ionized during one cycle. The light output at any instant due to these donors therefore depends only on the voltage $V_0 \sin \omega t$ at that time, i.e. $B(t) = B(V_0 \sin \omega t)$, and then the average light output is given by

$$\bar{B} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} B(V_0 \sin \omega t)\, dt = \tfrac{1}{2} \int_0^{2\pi} B(V_0 \sin x)\, dx$$

and this contribution to the electroluminescence is therefore independent of frequency.

The total light output can thus be divided into two parts, the first due to donors of depth less than E_0, and proportional to frequency, and the second due to donors of greater depth and independent of frequency. The critical energy E_c clearly increases with voltage, so that at high voltages the first component will predominate and the electroluminescence will be proportional to frequency. At low voltages, on the other hand, the second component will be the more important and the light output will be almost independent of frequency. This is in agreement with the experimental results reported in § 3.

5.2. *The Correlation of Electroluminescence and Conductivity*

According to the theory of this section, electroluminescence depends on temperature only through the ratio ω/σ and this is very important, for it establishes a relationship between electroluminescence and conductivity. The latter is given by $\sigma = \sigma_0 \exp(-E/\boldsymbol{k}T)$ and if at frequencies ω_1 and ω_2 corresponding points of the electroluminescence–temperature curve occur at T_1° and T_2° respectively, then $\omega_1/\sigma(T_1) = \omega_2/\sigma(T_2)$ which enables E to be found. The value of E thus derived from the electroluminescence data has been found to agree closely with that obtained by direct measurements of the conductivity. This agreement gives strong support to the theory outlined above.

§ 6. CONCLUSIONS

Experiments on the electroluminescence of single crystals of ZnS have been discussed, and their bearing on possible mechanisms considered. It has been shown that the theories previously proposed (Piper and Williams 1954, Taylor 1954) cannot satisfactorily explain the effect of changes of temperature and frequency on electroluminescence, which are revealed by these experiments. Modifications have been proposed which increase the extent of agreement between theory and experiment.

A theory based on thermal emission of electrons from traps agrees with the data on the variation of electroluminescence with temperature only if it is assumed that there is a distribution of trap depths. This cannot be considered a satisfactory solution as the agreement is sensitive to the form of distribution assumed. Measurement of other properties of ZnS might establish the existence of such a trap distribution, but it would still be necessary to show that the traps so found were those effective in electroluminescence.

The theory presented in § 5 is more successful and accounts in a satisfactory manner for all the main effects observed in the experimental investigations. In particular it accounts for the temperature variation and many of the complicated frequency effects; it also predicts a relationship between electroluminescence and conductivity which has been verified by experiment. The support which this gives to the theory would be further strengthened if measurements of conductivity and electroluminescence were made simultaneously for the same crystal and contacts. An apparatus which will enable this to be done is being constructed.

To establish a fully satisfactory theory it would be necessary to carry out measurements on crystals containing controlled amounts of activators and impurities. It is difficult to grow crystals to meet this requirement, but the results already obtained suggest that measurements of the temperature variation of electroluminescence made on the phosphor in powder form would be valuable provided they are carried out in the absence of the usual dielectric binding material.

ACKNOWLEDGMENTS

The authors wish to thank Professor J. Sayers for placing the facilities of his department at their disposal, Dr. G. F. J. Garlick for suggesting the problem and for the loan of the ZnS crystals, and the Admiralty for their support and for permission to publish this paper.

REFERENCES

BURNS, L., 1953, *Trans. Electrochem. Soc.*, **100,** 572.
DESTRIAU, G., 1947, *Phil. Mag.*, **38,** 700.
GOBRECHT, H., HAHN, D., and GUMLICH, H. E., 1954, *Z. Phys.*, **136,** 623.
HALSTEAD, R. E., and KOLLER, L. R., 1954, *Phys. Rev.*, **93,** 349.
INCE, A. N., 1954, *Proc. Phys. Soc.* B, **67,** 870.
PIPER, W. W., and WILLIAMS, F. E., 1952, *Phys. Rev.*, **87,** 151 ; 1954, *Brit. J. Appl. Phys.*, Supplement 4, 39.
ROBERTS, S., 1952, *J. Opt. Soc. Amer.*, **42,** 850.
TAYLOR, J. B., 1954, *Brit. J. Appl. Phys.*, Supplement 4, 45.
WAYMOUTH, J. F., 1953, *Trans. Electrochem. Soc.*, **100,** 81.
WAYMOUTH, J. F., and BITTER, F., 1954, *Phys. Rev.*, **95,** 941.
WAYMOUTH, J. F., JEROME, C. W., and GUNGLE, W. C., 1952, *Sylvan. Tech.*, **5,** No. 3.
ZALM, P., DIEMER, G., and KLASENS, H. A., 1954, *Philips Res. Rep.*, **9,** 81.

RESEARCH NOTES

On the Jet Formation by Explosives with Lined Hemispherical Cavities

By SAMPOORAN SINGH
Defence Science Laboratory, Ministry of Defence, New Delhi, India

Communicated by H. Kolsky; *MS. received 27th April* 1955

THE X-ray flash photography (Evans 1950) of explosives with lined hemispherical cavities indicated that the liner progressively turns 'inside out' during collapse and squirts out a jet along the axis. Kolsky (1949) also studied the deformation of a hemispherical liner by interposing a water layer between the explosive and the metal, and suggested that a rather different mechanism is operative here as compared with a conical liner. The present paper is an attempt to explain these observations by extending the conventional hydrodynamic theory of jet formation by lined conical cavities (Birkhoff *et al.* 1948, Pugh *et al.* 1952).

Let us consider a hemispherical liner divided into small elements, each element subtending an angle α† with the axis of propagation of detonation. The angle varies from 90° to 0° as a plane detonation wave sweeps from apex to base. As the pressures of detonation are very high compared with strength of metals, each element can be treated as a perfect fluid. The individual zonal elements are considered independent of one another and all the relations needed are obtained by applying the laws of conservation of mass, momentum and energy and Bernoulli's theorem to individual zonal elements.

Shreffler and Deal (1953) have shown that the velocity of a metal plate of thickness 0·06 in. to 0·1 in. driven by high explosives is approximately one half of the velocity of detonation. The usual thickness of liners met in shaped charges of 2 in. to 3 in. calibre are about 0·06 in. to 0·1 in. and it may be assumed that the collapse velocity of an element of liner, when $\alpha = 90°$, is about one half the detonation velocity. The experiments on 'inverted' 80° conical liners (Singh 1955) indicate that the element near the apex of the liner collapses with a velocity of 3100 m sec^{-1}. Eichelberger and Pugh (1952) find in the case of a 44° steel conical liner a velocity of 2850 m sec^{-1}. Walsh, Shreffler and Willig (1953), in their wedge-shaped collision experiments, showed that the plate (mild steel) velocity at 30° collision is 2520 m sec^{-1}. The above experiments lead us to assume an approximate expression connecting velocity of collapse V_0 and α as follows :

$$V_0 = \tfrac{1}{4} U_{\mathrm{d}}(1 + \sin\alpha) \qquad \ldots\ldots(1)$$

where U_{d} is the velocity of detonation of the explosive. U_{d} is constant as the detonation wave travels from the apex of the liner to the thin belt of explosive around the base. V_0 decreases continuously from apex to base, but each V_0 is assumed to be independent of the velocities of the neighbouring elements. The masses of the ring-shaped elements of the hemispherical liner increase from apex to base and near the base the belt of explosive surrounding the liner is much thinner; this results in further decrease of collapse velocities† from apex to base but, as a first approximation, this is neglected in the present paper.

---

† The angle between the detonation wave front and normal to the small element is also α.

Consider a system of coordinates moving with constant velocity $U_d \sec \alpha$ (a small element is considered where α is constant): in this system the detonation wave is stationary but the given zonal element would enter the detonation wave front with a velocity $U_d \sec \alpha$. Since the pressures arising from the detonation wave are everywhere perpendicular to the motion of the element, there results only a change in direction, and not in magnitude, of the element velocity. The arguments given by Pugh *et al.* (1952) are then valid and it can be shown that

$$V_0 = 2U_d \sec\alpha \sin\delta \qquad \ldots\ldots(2)$$

where δ is the angle between the direction of V_0 and the perpendicular to the original liner element. V_0 remains constant until the axis is reached. Combining equations (1) and (2), we have

$$\delta = \sin^{-1}\left(\tfrac{1}{8}\cos\alpha + \tfrac{1}{16}\sin 2\alpha\right). \qquad \ldots\ldots(3)$$

The position of P in the parent liner (figure 1) is fixed by a length M measured from the apex along the axis to the plane of the zonal element P. After collapsing,

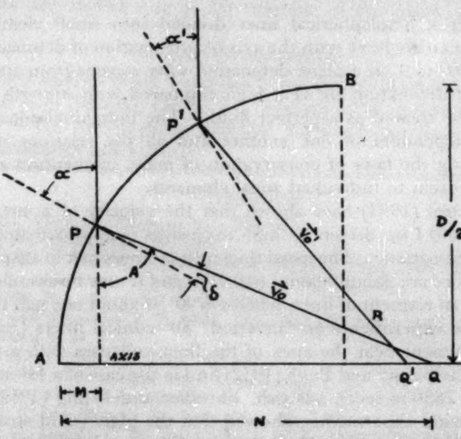


Figure 1. APB is the upper half and AQ the axis of the original hemispherical liner. When a detonation wave sweeps from apex to base along the liner, the elements P and P′ move towards the axis with velocities V_0 and V_0' and reach the axis at Q and Q′ respectively.

the element P reaches the axis at Q and the position of Q is fixed by a length N measured from the apex along the axis. Let $T(=M/U_d)$ represent the time the detonation wave takes in travelling from the apex to the liner element P, and $t-T$ the time element P takes to travel to Q. The distance travelled by P in the direction PQ at any time $\bar{t}$ is $(\bar{t}-T)V_0$. The distances travelled by different elements of the liner can be calculated at any time $\bar{t}$ and the shape of liner can be graphically traced.

---

† This is analogous to the collapse velocity gradient in the case of conical liner (Eichelberger and Pugh 1952), where the collapse velocities for an element near the apex and an element near the base of a conical liner are 2850 m sec^{-1} and 1500 m sec^{-1} respectively.

Figure 2 shows the stages of hemispherical liner collapse (taking $U_d = 7510$ m sec^{-1}) at different intervals of time. The shape E shows that the top (apex) of the hemisphere is forced forwards until in F the liner is turned completely 'inside out' into the form of a cone with a rounded apex. In shape G, the liner can be seen to have folded up and the jet has started travelling to the right. The apex of the liner forms the head of jet, while the sides fold back and follow behind. The predicted shapes† shown in figure 2 conform very closely to those observed by Kolsky and with x-ray flash photography.

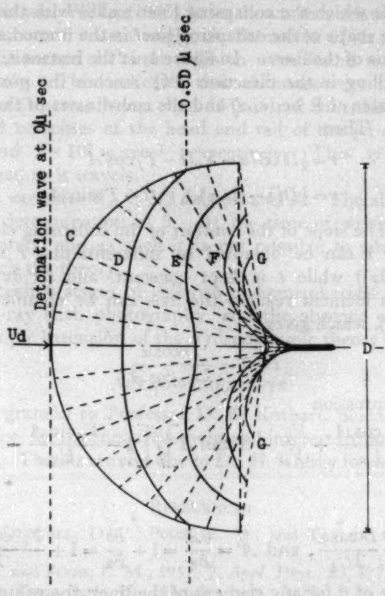


Figure 2. Stages of hemispherical liner collapse at different intervals of time. The liner is divided into 18 equal segments, the position of each element is calculated at different intervals of time and the shapes of the liner are graphically traced. The shapes C, D, E, F and G are at 0 μsec, 0·5D μsec, 1·0D μsec, 1·5D μsec and 2·0D μsec respectively.

When an element of mass dm reaches the axis, it splits up into two parts, one part of mass dm_j going into the jet and the other part of mass dm_s into the slug;

---

† This progressive turning 'inside out' of the liner becomes more pronounced if the decrease in collapse velocities from apex to base of the liner, due to increasing masses of the ring-shaped elements and decreasing thickness of the explosive, is taken into consideration.

these proceed along the axis at the constant velocities V_j and V_s respectively. The expressions for V_j and dm_s can be derived by arguments given by Pugh *et al.* (1952) and are as follows:

$$V_j = V_0 \operatorname{cosec} \tfrac{1}{2}\beta \cos(\alpha + \delta - \tfrac{1}{2}\beta) \qquad \ldots\ldots(4)$$

$$dm_s = dm \cos^2 \tfrac{1}{2}\beta \qquad \ldots\ldots(5)$$

where β is the angle which the collapsing liner makes with the axis. The angle β depends upon the shape of the collapsing liner in the immediate neighbourhood where it cuts the axis of the liner. In figure 1, at the instant t, P reaches the axis at Q, and P′ travelling in the direction P′Q′ reaches the position R. Let the cylindrical coordinates of R be (r, z) and the coordinates of the original position of P $(\tfrac{1}{2}D\cos\alpha, M)$. Then

$$r = \tfrac{1}{2}D\cos\alpha - V_0(t-T)\cos A \qquad \ldots\ldots(6)$$

$$z = \tfrac{1}{2}D(1-\sin\alpha) + V_0(t-T)\sin A \qquad \ldots\ldots(7)$$

where $A = \alpha + \delta$. The slope of the contour of the collapsing element of the liner at any given time t can be obtained by differentiating r with respect to z $(\partial r/\partial z = (\partial r/\partial\alpha)(\partial\alpha/\partial z))$ while t is kept constant, and $(\partial r/\partial z)_{r=0} = \tan\beta$. The time when a given element reaches the axis can be obtained by substituting $r = 0$ in equation (6), which gives

$$t - T = \frac{D\cos\alpha}{2V_0\cos A}. \qquad \ldots\ldots(8)$$

β is given by the expression

$$\tan\beta = \left[\tan\alpha + \frac{V_0\cos A}{U_d} + \frac{V_0'}{V_0} - A'\tan A\right] \Bigg/ \left[1 - \frac{V_0\sin A}{U_d} - \frac{V_0\tan A}{V_0'} - A'\right] \qquad \ldots(9)$$

where

$$V_0' = \frac{\partial V_0}{\partial\alpha}, \ \frac{\partial T}{\partial\alpha} = -\frac{D\cos\alpha}{2U_d}, \text{ and } A' = \frac{\partial A}{\partial\alpha} = 1 + \frac{\partial\delta}{\partial\alpha} = 1 + \frac{V_0'\cos\alpha - V_0\sin\alpha}{2U_d\cos\delta}.$$

To obtain the value of β for any element of the liner, the values of α, A, V_0, and U_d are substituted in equation (9). The value of β decreases continuously from apex to base of the liner; it is 170° for an element at the apex of the liner and 120° for an element at the base. It was simpler, however, to substitute values of α, A, V_0 and $t - T$ into the parametric equations (6) and (7) from which equation (9) was derived and to plot the resulting values of z and r for various instants of time. By drawing separate curves through the points determined for each instant of time, a series of contours of the collapsing liner were obtained. The angle between a contour and the axis at the point of intersection was the value of β for that particular element.

The distribution of mass dm into dm_j and dm_s is determined by β only. The major part of the metal of the element goes into the jet and so the whole jet would appear to be wider compared with the fine (narrower) jet from a conical liner. Obviously, the physical properties of metal control the characteristics of jets from hemispherical liners to a greater extent than those from conical liners. The whole hemispherical liner is progressively used up in the formation of jet and there is no massive slug. The absence of massive slug was confirmed by us by firing the liners in a tank of water.

In the notation of figure 1, $N' < N$ and the slug-element S from element P is ahead of the jet-element $J_{P'}$ from element P′. Thus since the slug from any element is ahead of the jet from a subsequent element it gains extra momentum at the expense of jet due to this subsequent element. The whole jet, which travels axially, resembles a long, thin, probably discontinuous tube and the formation of the jet and slug is a continuous process.

Walsh, Shreffler and Willig (1953) have shown that when the angle between a wedge is below a certain critical angle, the formation of jet does not take place. The 25° iron collision and 30° iron collision were found to be jetless and jet-forming respectively. This indicates that elements very near the base of a hemispherical liner (where α is less than 14°) do not contribute to jet formation.

The calculated velocities of the head and tail of the jet are of the order of 4×10^3 m sec^{-1} and 2×10^3 m sec^{-1} respectively. This gradient in velocity causes it to lengthen as it travels.

There are five variables V_0, δ, V_j, dm_s/dm and β. The absence of slug rules out experimental determination of β. At the time of writing, only V_j can be determined accurately, and as such it is not possible to advance experimental evidence † to check the theory.

Thus the successive stages of collapse of hemispherical liners observed by Kolsky and by x-ray flash photography, and the absence of massive slug are explained by a simple extension of the hydrodynamic theory of jet formation by conical liners.

ACKNOWLEDGMENTS

The author is grateful to Professor D. S. Kothari, Scientific Adviser to the Ministry of Defence, for his encouragement and interest throughout the course of this investigation. Thanks are also due to Dr. H. Kolsky for valuable suggestions.

REFERENCES

BIRKHOFF, G., MACDOUGALL, D. P., PUGH, E. M., and TAYLOR, G. I., 1948, *J. Appl. Phys.*, **19**, 563.

EICHELBERGER, R. J., and PUGH, E. M., 1952, *J. Appl. Phys.*, **23**, 537.

EVANS, W. M., 1950, *Trans. Instn Min. Metall., Lond.*, **2**, 17.

KOLSKY, H., 1949, *Research, Lond.*, **2**, 96.

PUGH, E. M., EICHELBERGER, R. J., and ROSTOKER, N., 1952, *J. Appl. Phys.*, **23**, 532.

SHREFFLER, R. G., and DEAL, W. E., 1953, *J. Appl. Phys.*, **24**, 44.

SINGH, S., 1955, *Proc. Nat. Inst. Sci. India*, **21**, 129.

WALSH, J. M., SHREFFLER, R. G., and WILLIG, F. J., 1953, *J. Appl. Phys.*, **24**, 349.

---

† Eichelberger and Pugh (1952) presented experimental verification of the theory of jet formation by conical liners by determining V_j and dm_s/dm for each element of the liner.

Isothermal Pressure Dependence of the Dielectric Properties of Eugenol, Glycerol and Water†

By B. K. P. SCAIFE‡
Department of Electrical Engineering, Queen Mary College, London

MS. received 15*th April* 1955

This note gives a brief account of some experimental work on the effect of high hydrostatic pressure on the dielectric properties of eugenol, glycerol and water.

The design of the high pressure apparatus used in these experiments follows that of Bridgman (1949). The maximum pressure obtainable was 12 000 kg cm^{-2}. The high pressure cylinder was pre-stressed by an autofrettage technique due to Manning (1945, 1950). Wave rings also due to Manning (Newitt 1940), were used to seal the high pressure joints of the apparatus. Bridgman's (1949) method for taking an insulated lead into the high pressure cylinder for electrical measurements was adopted. In this case the insulant used was sapphire in place of mica. The pressure in the high-pressure cylinder was inferred from a knowledge of the pressure in the low-pressure cylinder. It was possible to estimate the pressure in the high-pressure cylinder, after allowance had been made for friction of the piston packings, to within ±5%.

A radio-frequency bridge was employed for the electrical measurements. The particular bridge used was developed by Mayo, Maurice and Minns (Kirke 1945). With this bridge measurements of the dielectric constant and dielectric loss could be made in the frequency range $10^4-5\times10^6$ c/s. The accuracies within which it was possible to make capacitance and resistance measurements were ±2% and ±5% respectively. The accuracies of the derived results for ϵ' and ϵ'' (real and imaginary parts of the complex permittivity) were within ±4% and ±10% for substances with small values of ϵ' ($\epsilon'<10$). As ϵ' increased, so the accuracy of the value derived from experimental data increased.

In the figure ϵ'' for eugenol is plotted as a function of pressure for several frequencies. It will be seen that the frequency at which maximum absorption occurs decreases with increasing pressure. Table 1 contains the values of $\ln\tau$ for various pressures measured at a temperature of 20°C; τ is defined by the relation $\tau^{-1}=2\pi\times$ frequency at which $\epsilon''_{\max}$ occurs. The results for eugenol are not of sufficient accuracy to permit a detailed analysis of (ϵ'', pressure) curves, but it will be seen from table 1 that $\ln\tau$ is approximately a linear function of pressure.

Table 1

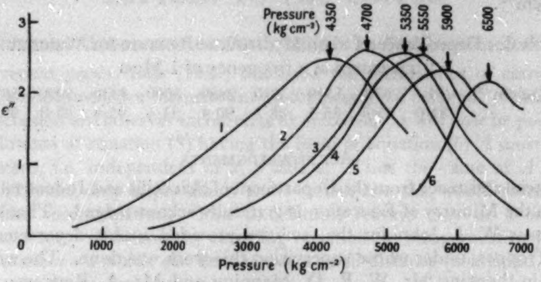
Pressure (kg cm^{-2})	4350	4700	5350	5550	5900	6500
$\ln\tau$	−17·24	−15·97	−15·35	−14·94	−14·25	−13·34

---

† The work described here forms part of a thesis approved by the University of London for the Ph.D. degree.

‡ Now at School of Theoretical Physics, Dublin Institute for Advanced Studies.

The values of ϵ' and ϵ'' at various pressures, for glycerol, are given in table 2. Measurements were made at two frequencies. The temperature was 5°c in both cases.



Isothermal pressure dependence of ϵ'' for eugenol at 20°c.

Curves: 1, 5·0 Mc/s; 2, 1·4 Mc/s; 3, 750 kc/s; 4, 500 kc/s; 5, 250 kc/s; 6, 100 kc/s.

Table 2. Dependence of ϵ' and ϵ'' on Hydrostatic Pressure for Glycerol at 5°c

Frequency 5 Mc/s			Frequency 1 Mc/s		
Pressure (kg cm⁻²)	ϵ'	ϵ''	Pressure (kg cm⁻²)	ϵ'	ϵ''
1	48·6	7·31	1	49·3	1·59
344	48·8	7·86	700	50·2	2·80
1030	48·7	9·1	1430	51·2	2·80
1350	48·1	10·5	2100	51·9	3·34
1700	47·7	11·9	2720	52·8	4·96
2050	46·8	13·1	3420	52·9	5·71
2400	45·8	15·6	4070	53·2	7·35
2720	44·6	18·1	4730	52·6	9·75
3030	43·2	19·2	5040	52·3	11·9
3330	41·7	22·1	6070	50·6	14·2
3440	40·0	22·8	6700	47·8	17·1
4000	37·8	23·4	7300	45·8	18·4
4340	35·4	24·7	7850	41·3	18·7
4700	33·2	25·5	8500	39·6	20·4
5000	30·9	26·3	8990	36·8	20·8
5300	28·8	26·3	9170	35·9	20·4
5470	26·8	26·3	9500	33·6	19·9
6050	25·0	25·4			
6980	18·8	22·7			
7700	15·2	20·3			
8000	14·7	19·2			
8500	14·1	18·1			

$\epsilon''_{\max} = 20{\cdot}8$ at 8800 kg cm⁻² at 1 Mc/s (from graph)

$\epsilon''_{\max} = 26{\cdot}3$ at 5300 kg cm⁻² at 5 Mc/s (from graph)

The dielectric constant ϵ' of water was measured as a function of pressure at a frequency of 1 Mc/s and at a temperature of 20°c.

The results for eugenol and glycerol confirm and extend the earlier important work of Danforth (1931). He found that eugenol, glycerol and iso-butyl alcohol exhibited the phenomenon of anomalous dispersion at high pressure. Danforth confined himself to measurements of ϵ' only.

The pressure dependence of ϵ' for water is indicated in table 3. The values of ϵ' for pressures below 3000 kg cm^{-2} are in good agreement with those of Kyropoulos (1926) who had made similar measurements up to a pressure of 3000 kg cm^{-2}.

Table 3. Dependence of ϵ' on Hydrostatic Pressure for Water at 20°c, measured at a frequency of 1 Mc/s

Pressure (kg cm^{-2})	1	670	1340	2010	2680	3660	4330	5000	5670	6000
ϵ'	81·0	83·4	85·5	88·7	90·4	93·4	95·4	97·0	99·0	100·0

ACKNOWLEDGMENTS

Financial assistance from the Department of Scientific and Industrial Research and from the Ministry of Education is gratefully acknowledged. Thanks are due to Professor W. J. John for the facilities provided in his department and to Dr. H. Tropper, under whose supervision this work was done. The writer takes pleasure in thanking Mr. W. R. D. Manning and Mr. A. Kennaway (Plastics Division, I.C.I.) for their invaluable advice concerning the design and construction of the high pressure apparatus.

REFERENCES

BRIDGMAN, P. W., 1949, *The Physics of High Pressure* (London : Bell).

DANFORTH, W. E., 1931, *Phys. Rev.*, **38**, 1224.

KIRKE, H. L., 1945, *J. Instn Elect. Engrs*, Pt III, **92**, 2.

KYROPOULOS, S., 1926, *Z. Phys.*, **40**, 507.

MANNING, W. R. D., 1945, *Engineering, Lond.*, **159**, 101, 183; 1950, *Ibid.*, **169**, 479, 509, 562.

NEWITT, D. M., 1940, *The Design of High Pressure Plant and the Properties of Fluids at High Pressures* (Oxford : University Press).

LETTERS TO THE EDITOR

Carrier Concentration Disturbances in Semiconductors

In a recent paper, Low (1955) discussed the disturbance of carrier distribution in a semiconductor filament during the passage of current. Some features of the discussion are however incomplete or erroneous as will now be pointed out.

In solutions of equation (5) having the form of equation (6) A must be taken as a constant, i.e. independent of x, y and z. Thus the value of A given by equation (11) can only imply that b and c vanish and this would correspond to zero surface recombination. This undesirable feature may be avoided in a number of ways, for instance by retaining equation (6) but permitting p_1 to vary over the plane $x=0$ which can be taken at a small distance (compared with $1/a$) from the actual metallic electrode.

It was stated at the outset that the filament was assumed long and this was the reason for taking only the positive root in equation (8). In this case the term in $\exp(-aL)$ in equations (13), (16) and (17) must be dropped for consistency. This may be seen when the exact analysis is carried out for a filament of arbitrary length which requires retention of both roots in the solution. It is then found that the total increment in the number of holes in the filament is simply given by

$$P_1=(\gamma_1-\gamma_2)\tau_f I/e \qquad \ldots\ldots(1)$$

where γ_1 and γ_2 are the hole injection and extraction ratios at the positive and negative electrodes respectively.

In Low's equation (16) it may be noted that the denominator is simply e/τ_f without approximation. Thus the provisos following equation (17) and the requirement of negligible diffusion mentioned in §4 are unnecessary restrictions.

In experimental work on filaments with only two electrodes (e.g. Many 1954) it is sometimes overlooked that the perturbations of carrier distribution and of conductance depend on the effects at *both* electrodes since $(\gamma_1-\gamma_2)$ is involved. This effect is particularly important when weakly injecting or extracting contacts are being studied. The dependence of the change of conductance on current is seldom linear since $(\gamma_1-\gamma_2)$ is itself usually a function of current ; however accurate bridge measurements, say to 1 part in 10^4 or better, will reveal the closeness with which non-ohmic contacts have been achieved when these are required in studies of current noise and photoconductivity.

Department of Physics,
University of British Columbia,
Vancouver 8,
Canada.
18th July 1955.

R. E. Burgess.

Low, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 310.
Many, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.

REVIEWS OF BOOKS

Les Fonctions de Bessel (2nd Edn), by G. GOUDET. Pp. 90. (Paris: Masson et Cie, 1954.) 600 *fr.*

This is an introductory text, which is addressed particularly to physicists. The first three chapters are devoted to the introduction of Bessel functions as solutions of Bessel's equation together with most of the better known recurrence formulae, integrals and expansions. The final chapter gives an account of some physical applications of the foregoing. These cover the oscillations of a cylindrical cavity, the skin effect, conduction of heat by a cylinder, and finally a group of problems in diffraction. The treatment is simple and a model of clarity. The inclusion of examples, however, would have been an advantage for the student. H. H. HOPKINS.

Laplace Transforms for Electrical Engineers, by B. J. STARKEY. Pp. 279. (London: Iliffe, 1954.) 30*s.*

This book is intended as an introduction to its subject for electrical engineers whose mathematical knowledge is limited to " the differential and integral calculus, and the principles of the theory of complex numbers and of vector analysis ". The subject is amply covered, with many well-chosen illustrative examples, up to the use of the Inversion Theorem for single-valued transforms. The topics discussed include the use of partial fractions for inversion, stability criteria, pulses and step-functions, the Product and Duhamel Theorems, and the transforms of current and voltage in a general transmission line. There are two very useful tables of transforms, with nearly 200 entries, and an interesting 24-page collection of formulae and references to tables of various ' higher ' functions: gamma, Bessel, etc.

What distinguishes this book from others on the same subject is its novel treatment of analytic functions and contour integration. Differentiation and the Cauchy–Riemann equations appear only in an Appendix, for, says the author (p. 101) " the role played by differentiation " (of analytic functions, in physical and engineering problems) " is quite insignificant ", (!), " whereas integration is of paramount importance ". An outline of the route followed may therefore be of interest.

Briefly, the function p^n, for various n, is integrated along (a) arcs of circles around the origin, (b) segments of rays issuing from the origin, and hence round any contour by polygonization into elements of types (a) and (b). The different types of functions and singularities illustrated by p^n for different n are discussed and classified according to the values of integrals on closed contours including or excluding the origin. Functions behaving similarly under integration are described as analytic, and are stated (naturally without proof) to be always expressible as combinations of powers $(p-a)^n$, with various a and n. Maclaurin's and Laurent's series, and eventually the Residue Theorem, are thus obtained.

There is of course no attempt at rigour here, nor is this necessary for the purpose of the book. I fear, however, that the student will find in the form of presentation of the above and subsequent theory (as distinct from the illustrative practical examples) anything but the compensating simplicity claimed by the author in his Introduction. An orthodox treatment, with fewer improved statements, could have been given in fewer pages and, moreover, would have led on naturally to more advanced texts—which is hardly the case with this book. For example, instead of a simple explanation of how the Bromwich contour (i.e. line from $c-j\infty$ to $c+j\infty$) can be closed, for the purpose of applying the Residue Theorem, by addition of an infinite semicircle to the right or left, there is a somewhat mystical and unconvincing appeal (p. 146) to a global model of the complex plane, in which one pole represents the origin and the other the point at infinity. This leads to the remarkable statement that " the result of integrating a meromorphic function along a closed contour extending to infinity, which can be represented as a line extending to infinity in both directions and dividing the complex plane into two infinite parts, is the same for both parts of the plane apart from sign ". This is followed up in a similarly mystical way (pp. 150, 151) in proving the Inversion Theorem.

Again, on p. 213, in discussing the integral of p^m for non-integral m on a Bromwich contour (ACB), the spherical model leads, mysteriously, to an added contour on either side of the cut negative real axis (and, incidentally, running *through* the origin instead of around it !). Thereupon the original contour is discarded since " the integral along ACB is zero "—which of course is nonsense. The author's whole object, throughout, seems to be to avoid discussing the behaviour of functions at infinity. Indeed, at the only place where he ever mentions such behaviour (p. 143), he states, quite wrongly, that an integral on a large semicircle tends to zero if the integrand does so.

Some further errors must be briefly mentioned. The factor π or 2π is omitted in a dozen of the formulae (pp. 22–27) on Fourier series and integrals. Eqn. (1.40) implies that $e^{-j\omega t}=0$ for $t=\infty$. A totally wrong factorization, as if for a polynomial, of $1\pm e^{-ap}$ is given (unnecessarily) on p. 193. The transform of t^n (p. 159) only exists if $n>-1$, but this is nowhere pointed out. The author does not realize (pp. 146, 188) that the conventional (and mis-named) ' residue at infinity ' is not necessarily associated with a pole at infinity. (Thus, the function $1/p$ has ' residue ' -1 at infinity, but no pole there; p has ' residue ' 0, yet has a pole.) The principal value of the integral of p^n, $n<0$, on a contour through the origin is stated (p. 116) to be ∞; in fact it is 0 if n is odd $\neq -1$, or if $-1<n<0$. A very strange discussion (pp. 29, 30, 144) leads to the result that the real integral

$$\int_{-x_0}^{x_0} \frac{\cos x}{x}\,dx$$

has the value $j\pi$!

The book contains some good and useful sections, and it might have been well worthy of recommendation, if only the author had been able to distinguish between lack of rigour and sheer inaccuracy, and had not, in places, used metaphysics as a substitute for mathematics. A. TALBOT.

Introduction to Mathematical Statistics, by P. G. HOEL. Pp. xi + 331. (2nd Edition.) (New York: John Wiley; London: Chapman and Hall, 1954.) 40*s*.

The satisfaction with which the first edition was received has encouraged Dr. Hoel to extend the contents of his book. The result is, for the student with an appreciation of lucid and moderated mathematics in a context of commonsense discussion and illustration, one of the best introductions to statistics now available.

The chapter on probability (which is now included!) provides the very necessary foundation for the whole subject and manages to be both compact and clear. It is immediately followed by a discussion of the main principles involved in the testing of hypotheses and the estimation of parameters, bringing the reader to the heart of the subject, so that he knows his purpose before studying the means of achieving it.

The computation of descriptive measures for empirical distributions of one variable leads into an analysis of the properties of the common model distributions by immediate use of the moment generating function, rounded off by illustrations of the application of the mathematical models to simple practical problems with ample warning that only rarely will the skeleton model appear as a perfect representation of reality.

After a concise explanation of the basis of sampling theory for one variable (large samples) and a descriptive and theoretical treatment of calculation and regression, the application of χ^2 for testing goodness of fit rounds off the first half of the book. Dr. Hoel now leads the reader to a more critical examination of the principles involved in testing simple and composite hypotheses: the Neyman–Pearson lemma of critical regions and the use of the likelihood ratio. This is followed by a more sophisticated account of the theory of estimation and a discussion of small sample theory to provide the necessary understanding of such significance tests as are normally met with in the analysis of experimental data.

It is now taken for granted that any book on statistics should not only stress the importance of good quality experimental data but also tell the reader something of the art of collecting it. Dr. Hoel sketches very briefly general principles for the design of experiments in which components of variation can be separated by the method of analysis of variance, stressing the link between a hypothetical model and the numerical analysis of the data. He regards industrial sampling inspection as essentially a branch of the design of experiments directed to obtaining desired quality production at nominal cost with specified risks.

The final chapter deals with the now very fashionable 'non-parametric' or 'distribution-free' methods, which our American friends more accurately and more realistically call 'quick and dirty'.

The whole book is attractively done and can be strongly recommended to those who are looking for a nice beginning that will carry them a very long way in practice. Dr. Hoel leaves no excuse for narrowness: every chapter has a list of references. Equally, he leaves no room for complacency: every chapter has an extensive set of exercises. But why no answers?

M. J. MORONEY.

Introduction to Experimental Physics, by W. B. FRETTER. Pp. viii + 349. (London: Blackie and Son Ltd., 1955.) 35*s*.

According to the author's Preface this book has been compiled from notes taken at a course of lectures in experimental techniques given for some years to graduate students at the University of California. It is therefore not surprising to find that the techniques chosen for description are treated rather sketchily and in many cases inadequately from the point of view of the prospective user. The detailed instructions with which one had become familiar in books such as Strong's *Procedures in Experimental Physics* are given considerably less prominence in the present volume. On the other hand the condensed presentation and the shortness of the chapters makes for easy reading. On reaching the end of the book the reader may well be left with the feeling that he has accompanied Professor Fretter on a tour of the Physical Laboratories of the University of California. He will probably retain after a single reading the same sort of impressions that might be gathered from such a tour.

Most of the 32 chapters of the book are devoted to atomic and nuclear physics although many of the techniques described are useful in other fields. Thus the first chapters deal with selected topics in electrical and magnetic measurements, electronics, magnet design and vacuum techniques. The next group of chapters gives a condensed account of methods of particle and radiation detection. The material presented is obviously incomplete and is chosen mainly to acquaint the reader with the ' vocabulary ' of the subject and to suggest further reading. Chapters 16–20 are mainly descriptions of the large accelerating machines at Berkeley. Finally there are sections on cosmic-ray techniques, mass spectroscopy, β-ray, optical and microwave spectroscopy, molecular beams, magnetic resonance, pile techniques, x-ray diffraction and low temperature physics (with an average of about 12 pages each). The book concludes with a chapter on laboratory hazards (radiation exposure and toxicity of chemicals) and an interesting philosophical section on the design of experiments, which might well have prefaced the volume.

Professor Fretter has achieved his object of presenting a wide range of topics in readable form. The book is full of interesting pieces of information and is likely to remind most readers of some long-forgotten fact or trick as well as to suggest some new back-of-the-envelope calculation. Those who are giving or planning graduate courses on techniques will find the book useful both for its method of presentation and for the adequate but not overwhelming bibliographies at the end of each chapter. Altogether the author may be congratulated on having produced a volume which will both help and stimulate those who profess some interest in the arts of atomic and nuclear science.

W. E. BURCHAM.

The Students' Physics, Vol. I—Light, by R. W. DITCHBURN. Pp. 680. (London and Glasgow : Blackie, 1952.) 45*s*.

This new volume of a well-known series is already, in a short time, an established part of the honours physics literature. Prior to the appearance of Professor Ditchburn's admirable book there was an acute difficulty in recommending to students a reference work of the required standard. The shelves of bookshops abounded with books suitable for a first-year introductory course, but none of sufficient scope, modernity, and standard to satisfy the needs of a full honours course. In consequence teachers of optics were at a disadvantage. This difficulty has now been removed by the publication of an adult account of optics suitable for honours students, and the subject has been put in its proper setting as a progenitor of present-day wave physics.

Professor Ditchburn's book is presented in nineteen chapters, having in most cases a set of examples, references and appendices dealing with more advanced topics. The first three chapters comprise a historical introduction, and a full introduction to the theory and properties of waves. Of real importance to the student is the discussion and analysis of wave trains, which, together with a digest of the necessary parts of the Fourier analysis, makes up chapter IV. The treatment of group velocity could be recommended not only to students, but to many teachers, judging by the number of books that calculate the velocity of a ' beat ' envelope without relating it to the velocity of a more complex group.

Chapters V and VI deal with the principles of interference and diffraction and with the theory of a number of experimental devices, the uses of interferometers being reserved for chapter IX. The derivation of Kirchhoff's integral is given in an appendix. Further discussion of Huygens' principle and Fresnel diffraction phenomena is given in chapter VII, together with the connections between ray and wave optics. The following two chapters take up the question of resolution in optical instruments, and chromatic resolving power of spectroscopic devices, and the applications of interferometers. There follow chapters on the velocity of light, relativity and polarized light.

It is Professor Ditchburn's avowed aim to treat the ' different ' theories of light as aspects of what he regards as " a single theory of light, a logical whole ". As part of this plan the electromagnetic theory of light is treated in some detail, including the theory of reflection, dispersion, absorption and scattering, together with relevant experiments. This is further developed in chapter XVI, dealing with anisotropic media.

In the last three chapters the more modern theories are treated. The first of these includes accounts of the photoelectric effect, the Rutherford–Bohr atom, and spectra, and the production of spectra. It is followed by an excellent introduction to wave mechanics and the theory of radiation. The final chapter outlines the quantum theory of the processes of emission and aborption.

There are a few mis-statements in the text, as is almost inevitable in a work of this extent. As an example, on page 170 it is stated that there is no simple relation between the illuminations or amplitudes at points of non-zero amplitude produced by complimentary diffraction screens, whereas a very simple rule does exist (Sommerfeld, *Optik*, p. 208). In certain places the mathematical presentation might be improved in minor ways, particularly in the important chapter on wave trains and Fourier analysis.

These slight criticisms, cannot, however, detract from the impressive achievement it has been Professor Ditchburn's good fortune to attain. Both the content and presentation leave, in general, nothing to be desired : the wealth of information and detail contained in the book is truly remarkable. The high standard of production of the book is uniform with that of the other volume of the series, and is well illustrated with diagrams and plates. One need not wish a book of this kind success—it will gain it, and in so doing will offer the possibility of a notable advance in the teaching of optics in honours physics courses.

H. H. HOPKINS.

Recent Advances in Optics, by E. H. LINFOOT. Pp. x + 286. (Oxford : Clarendon Press, 1955.) 50*s*.

This book contains descriptions of recent work, some unpublished, on several aspects of instrumental optics. The first section is on aberration balancing in optical systems, using a ray-theoretic treatment; the criterion chosen is that the effective image patch radius for a given object point shall be a minimum, this radius being defined as the radius of gyration of the density of spots in the spot diagram about the centroid. A complete balancing system for primary and secondary monochromatic aberration coefficients and gaussian and primary chromatic coefficients is given. The second section deals with diffraction effects in point images; a compact exposition is given of recent work by Nijboer and others on images formed in the presence of aberrations and with partially obscured pupils. The third section deals with imagery of extended objects; the effect of the degree of coherence of the illumination on the images of periodic structures is discussed and applications are given to the theory of the microscope.

The next section of the book is a re-presentation with extensions of the author's elegant and detailed analysis by diffraction theory of the Foucault knife-edge test. Detailed descriptions of the appearances to be expected with a variety of errors are given; these results should be of great value to the experimental optician.

The next section is a study of the Schmidt camera. The basic theory is given and then, taking account of second and fourth power terms in the corrector plate, the residual aberrations (extra-axial and chromatic) are calculated and displayed as spot diagrams. Next, higher order aberrations are introduced and the problem of aberration balancing by fourth power figuring on the concave mirror is discussed. Finally, a field flattening lens is introduced and the problem re-worked.

The last section of the book deals with plate-diagram analysis, which is a semi-graphical representation of the usual Seidel formulae for the effects of stop-shifts and aspheric figurings. It is used in this book to investigate the possibilities of two-mirror systems with corrector plates as astronomical telescopes; several systems of the Schmidt–Cassegrain type are discussed in detail.

The author is very careful throughout to indicate the orders of magnitude of error terms in aberration expressions, a precaution which is often neglected by applied opticians; in this respect the book is a joy to read. There are some instances where one could wish for a statement about physical approximations; for example scalar diffraction theory is used throughout with the usual restriction to small angular apertures, but it is applied without warning to the theory of microscope images in partially coherent light where cones with semi-angles up to 65 or 70 degrees can occur. Again it should be pointed out that the radius of gyration of the image patch as used in the first section, a concept which seems to be due to C. F. Gauss, has limited physical meaning since the radius of gyration of the Airy disc is infinite, and so *a fortiori* is that of an aberrated point image. Bibliographies are given at the end of each section; they are obviously not intended to be exhaustive, but it seems a pity that the work of Berek in 1926 on partial coherence was not mentioned since it anticipates a number of the ideas of van Cittert and later workers. More uniformity in notation would be desirable throughout the book, particularly in the coordinates used in diffraction theory, where the usual dimensionless diffraction coordinates are given different symbols and defined in terms of different symbols in different sections.

These minor criticisms apart, the book can be recommended as an excellent review of recent work in these fields. W. WEINSTEIN.

CONTENTS OF SECTION A

PAGE

Mr. B. A. CHARTRES and Mr. D. A. TIDMAN. The Scattering-in Correction . 841

Mrs. R. B. J. PALMER and Dr. H. A. B. SIMONS. The Variation of Track Length with Angle of Dip in Photographic Nuclear Research Emulsion . . . 852

Dr. B. S. CHANDRASEKHAR and Dr. K. MENDELSSOHN. Sub-Critical Flow in the Helium II Film 857

Dr. P. G. HARPER. Single Band Motion of Conduction Electrons in a Uniform Magnetic Field 874

Dr. P. G. HARPER. The General Motion of Conduction Electrons in a Uniform Magnetic Field, with Application to the Diamagnetism of Metals . 879

Mr. H. TREFALL. On the Positive Temperature Effect in the Cosmic Radiation and the μ–e Decay 893

Mr. A. D. BUCKINGHAM and Dr. J. A. POPLE. Theoretical Studies of the Kerr Effect: I—Deviations from a Linear Polarization Law 905

Mr. A. D. BUCKINGHAM. Theoretical Studies of the Kerr Effect: II—The Influence of Pressure 910

Research Notes:

- Dr. R. H. TREDGOLD. Anharmonicity and Monomolecular Reactions . . 920
- Mr. S. W. PEAT and Dr. M. A. S. ROSS. Radioactive Disintegration of ^{232}Th 923
- Mr. D. L. ALLAN. The Energy Distribution of Protons from the Reaction Cu(n, p)Ni at 14 MEV 925
- Mr. K. R. CHAPMAN, Dr. E. MATSUKAWA, Dr. P. H. ROSE and Prof. E. A. STEWARDSON. The Absolute Measurement of Differential Cross Sections for the Nuclear Scattering of Low Energy Relativistic Electrons . . 928
- Dr. A. KELLER. Phase Shifts in p–p Scattering 930
- Mr. C. A. HAYWOOD. An Approximation for the High Order Phase Shifts in the Elastic Scattering of Slow Protons in the Inert Gases . . . 932

Letters to the Editor:

- Dr. J. M. BAKER and Dr. B. BLEANEY. Hyperfine Structure of Praseodymium 936
- Dr. B. BLEANEY. Nuclear Moments of the Lanthanons from Paramagnetic Resonance 937
- Dr. D. F. BREWER, Mr. D. O. EDWARDS and Dr. K. MENDELSSOHN. The Entropy of Superfluid Helium 939
- Dr. W. D. ALLEN and Dr. A. T. G. FERGUSON. Elastic Scattering of 80 kev Neutrons 940
- Dr. M. A. S. ROSS, Mr. A. J. COCHRAN, Dr. J. HUGHES and Prof. N. FEATHER. Fluorescence Yields of the L Levels of Bi 942
- Mr. W. E. PERRY. Comparison of British and United States National Radium Standards 942
- Mr. S. M. SHAH. Energy Levels of Mu-Mesic Atoms 945

Reviews of Books 947

Contents of Section B 952

The Superconducting Bolometer as a Detector of Thermal Radiation from Low-Temperature Sources

By J. A. HULBERT and G. O. JONES
Department of Physics, Queen Mary College (University of London), London E.1

MS. received 17*th November* 1954 *and in revised form* 22*nd June* 1955

Abstract. A superconducting bolometer has been developed as a detector of the thermal radiation from low-temperature sources. With an area of 1 cm^2, it has a sensitivity of about 3×10^{-12} watt absorbed for a time-constant of 20 seconds. It has been used in preliminary observations of the optical properties of a graphite-coated surface at about 4°K, using radiation emitted from a source at a temperature between 9° and 17°K and assumed to lie mainly between 0·2 and 0·4 mm in wavelength.

The superconducting bolometer (Goetz 1939, Andrews 1947, Fuson 1948, 1949) depends for its action on the high temperature coefficient of resistance, in the region of its transition temperature, of the superconducting metal which forms the receiving element. This instrument has high temperature sensitivity and the inherent advantage that noise arising from thermal fluctuations is limited because the receiver is at a low temperature.

As a preliminary to studies at wavelengths between the normal infra-red and microwave regions, we have developed a superconducting bolometer suitable as a detector of the thermal radiation emitted by an approximately 'black' body in the neighbourhood of 10°K. At this temperature the maximum emission of energy per unit frequency range (as given by Wien's displacement law) occurs at a wavelength of about 0·5 mm. However, as implied in Stefan's law, the available flux of energy from such a source is very small, and the receiver must be developed to its maximum possible sensitivity.

As in the earliest superconducting bolometers (Andrews *et al.* 1941, 1942), the method of continuous irradiation has been employed. This allows direct indication of the output of the bolometer by a sensitive galvanometer without amplification. In order to avoid stray radiation both the receiver and the source of radiation have to be contained in an evacuated enclosure maintained at about 4°K. The temperature of the source can be adjusted independently as required.

Figure 1 shows the experimental assembly. The bolometer element A, consisting of a fine tin wire glued in a zig-zag pattern to the back of a 1 cm square of thin aluminium foil, is mounted on a frame B of nylon thread and fine constantan wire. The only other attachments to the receiving element are extended electrical leads C, cut from tinned lead foil. At the working temperature these are fully in the superconducting state so that no Joule heat is dissipated in them by the current through the bolometer element. They also provide a higher thermal resistance between the receiving element and the base D of its mount than they would in the normal state.

The ambient temperature of the bolometer, in the region of the transition temperature of tin (about 3·7°K), is controlled, first, by regulation of the pressure above liquid helium in the vessel E. A fine control of temperature to better than 10^{-4} deg is provided by the use of a carbon thermometer F, linked through a photoelectric amplifier to a thyratron circuit operating a heater G.

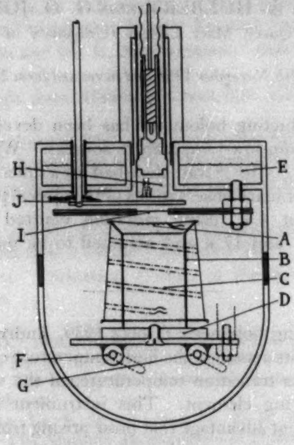


Figure 1. The superconducting bolometer and source of radiation.

The source of radiation consists of a series of Mendenhall wedges H. A thin window of Formvar I protects the receiving element from spurious effects due to residual gas, and a shutter J, which may be operated from outside the apparatus, is provided between radiator and detector. The whole assembly is fitted in the working space of the evacuated casing of a miniature helium liquefier (Chester and Jones 1953).

Changes in the potential difference across the receiving element, which carries constant current, are indicated by a galvanometer. Since the element is constructed to have good thermal equilibrium within itself under all operating conditions, its voltage sensitivity to absorbed power may be found by a direct calibration in terms of the electrical power derived from Joule heating in the absence of incident radiation. This method has the advantages that it does not require knowledge of the optical nature of the receiving surface and that non-linearity in the instrument may be found directly and corrected. In this way, the maximum useful sensitivity to absorbed power has been found to exceed 1500 μv erg^{-1} sec. With an ordinary sensitive galvanometer it is then possible to make measurements of absorbed power accurate to $\pm 3 \times 10^{-5}$ erg sec^{-1}. (The time constant of the bolometer as at present assembled is 20 seconds.) The performance is limited at present by fluctuations in the control of temperature, and is clearly capable of further improvement. However, the figures of performance just quoted represent advances over the corresponding figures as reported by Andrews by factors of about 10 and 30 respectively. The increased sensitivity appears to be due mainly to improved control of temperature and to the choice of tin, which has a very low heat capacity per unit volume near 4°K, instead of tantalum for the material of the receiving element. The other main improvement in the design of the present bolometer is that the whole of the superconducting element is affixed to the surface receiving radiation, making possible an accurate direct calibration by the method described above.

Figure 2 shows the results of some observations of the absorptivity of a thin layer of graphite deposited (using 'Aquadag') on the surface of the receiver,

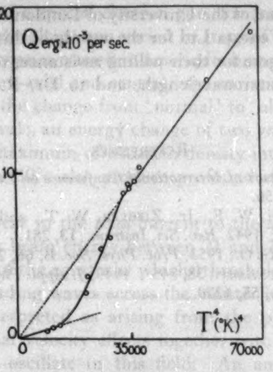


Figure 2. The power Q absorbed by a graphite-coated surface at 4°K plotted against the fourth power of the temperature T of the source of radiation. (The values represented cover the range 9° to 17°K.)

using a 'black' source of the type described in which the internal surfaces of the wedges also had been coated with graphite. The energy Q absorbed by the receiver has been plotted against T^4 (where T is the temperature of the source) in order that changes in the mean absorptivity should appear as departures from linearity. Making the working assumption that the absorptivity of a graphite surface is independent of temperature between 4° and 17°K and varies only with wavelength (and is therefore the same throughout for the receiving surface and for each internal surface of the wedges) the results would indicate a change in mean absorptivity of the receiver from about 2% to 4½% (the values which correspond to the slopes of the dotted lines) as the peak in the curve of energy distribution moved from about 0·4 mm wavelength to 0·2 mm. These low values of absorptivity imply that the system of Mendenhall wedges (each of 8° angle) used as the source of radiation can have been only about 30–60% black; this has been allowed for in making the above estimates, but without special consideration of the fact that the wavelength of maximum emission is here not much smaller than certain dimensions of the wedge system. It is also possible that a variation with *temperature* exists in the absorptivity of graphite, but the present results do not permit us to isolate this effect. We do not consider these observations as more than preliminary, but they are of interest as a reminder that low values of absorptivity are likely to be encountered in this region of wavelengths, as in the microwave region. (Although graphite is normally regarded as a good absorber in the infra-red, the absorptivity of a graphite-coated surface of the type used here was found to be less than ¼% at 8 mm wavelength and at room temperature.) For this reason we have in this paper everywhere specified sensitivities in terms of response to *absorbed* power rather than to incident power.

ACKNOWLEDGMENTS

We owe acknowledgments to the Department of Scientific and Industrial Research for a maintenance allowance to one of us and for a grant-in-aid, and to the Central Research Fund of the University of London for the loan of apparatus.

We are indebted to Venesta Ltd for the supply of thin foils, to Mr. W. A. G. Baldock and Mr. W. Eagers for their willing assistance, to Mr. R. F. Keating for the measurements at 8 mm wavelength, and to Dr. R. O. Davies for helpful discussions.

REFERENCES

ANDREWS, D. H., 1947, *Report of International Conference on Low Temperatures* (London : Physical Society), p. 56.

ANDREWS, D. H., BRUCKSCH, W. F., JR., ZIEGLER, W. T., and BLANCHARD, E. R., 1941, *Phys. Rev.*, **59,** 1046 ; 1942, *Rev. Sci. Instrum.*, **13,** 281.

CHESTER, P. F., and JONES, G. O., 1953, *Proc. Phys. Soc.* B, **66,** 296.

FUSON, N., 1948, *J. Opt. Soc. Amer.*, **38,** 845 ; 1949, *J. Appl. Phys.*, **20,** 59.

GOETZ, A., 1939, *Phys. Rev.*, **55,** 1270.

Some Anomalous Emission Effects in Low Energy Electron Beams

By W. W. H. CLARKE and L. JACOB
George Holt Physics Laboratory, University of Liverpool

MS. received 26th January 1955 *and in amended form 4th May* 1955

Abstract. Some of the properties of an electron emission system have been investigated by means of a special electron gun, for accelerating voltages in the range 18–30 volts. The voltage was varied in steps of 1 or 2 volt intervals, and it was found that the emission varied in a periodic manner over all regions of the cathode—marginal, extra-axial and central; the variations were greatest for the central region. Here, the change from 'normal' to 'abnormal' may occur over a one volt energy interval; an energy change of two volts can alter the periodic pattern and convert a maximum of emission density into a minimum. Changes of temperature of some 30°c also produce profound effects on the form of the distributions.

Anomalies also appear in the *total* current to the axial system of collectors. For example, for every beam energy between 21 and 23·8 volts, there are three separate temperatures of the emitter at which 'threshold' current begins to flow.

The pattern of standing waves across the emitter has a wavelength of about 0·01 mm. This is interpreted as arising from the presence of a saddle-back field produced by initial velocity effects together with the space charge of the beam. The electrons oscillate in this field. An analogy is afforded by the occurrence of oscillations under certain field conditions for the transition region between lens and mirror in the case of the symmetrical arrangement of electrodes.

§ 1. INTRODUCTION

In some work dealing with the characteristics and focusing of electron beams produced by a special type of electron gun in the energy range 20–70 volts, the authors (1951) recorded the presence of standing wave patterns in the electron distribution from the emitting cathode. This effect appeared to depend on the cathode temperature. Both 'hollow' and 'semi-hollow' beams were recorded in the distributions (that is, those in which the axial emission is practically zero, and less than that off the axis respectively). The remarkable changes of slope in the distribution curves for the 20–30 volt region appeared to be connected with changes of state in the cathode emission for particular fields, or with corresponding changes in charge density in the virtual cathode.

The formation of the standing wave patterns appeared to depend critically on changes in energy over a limited energy range of about ten volts. Outside this range, the electron distribution showed that beam formation was 'normal', and proceeded in a way to be expected from a formal analysis of the electron motion in the field, taking space charge into account.

The purpose of this paper is to present a detailed account of the experimental observations, particularly those which give rise to the anomalous distributions observed. Great care was taken to work over small energy intervals in the critical region and to ensure that the effects were real and readily repeatable.

§ 2. Apparatus and Method of Experiment

The form of the electron gun used in these experiments has already been described in an earlier paper (Clarke and Jacob 1953). It is reproduced in figure 1 in order the better to follow the subsequent results. It should of course be stated that the whole electrode assembly was in a sealed-off tube in which the pressure before bringing down the getter and sealing off was about 10^{-6} mm. All electrodes were hydrogen furnaced before assembly, and some further high-frequency heating was carried out on the accelerating system in the assembled unit on the pump.

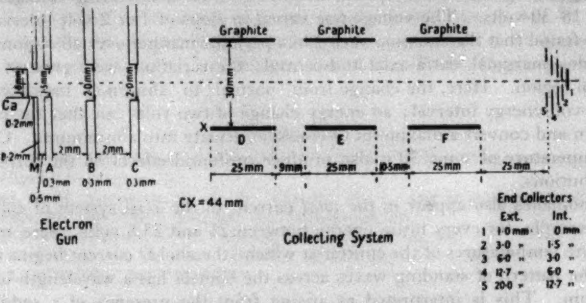


Figure 1. Immersion objective and collector system.

The dimensions and disposition of the set of concentric collectors at the end of the experimental tube which served the purpose of recording the beam distributions are also shown. The electrodes A, B, C, D, E, F and the collectors were all kept at the same potential (except for certain tests from time to time to check for the presence of secondary emission current) corresponding to the final energy of the beam; the currents collected by each of these in turn was recorded for various values of the modulation ratio α between 1·0 and 6·0; α is the ratio of the accelerating potential to the potential of the modulator with respect to the cathode. A further set for $\alpha = \infty$, that is, for the modulating electrode at cathode potential (which produced the largest beam angle) was also taken; both sets of curves can be considered as representative of the measurements.

When dealing with the effects of temperature on the emitting properties of the cathode, it was necessary to obtain reliable estimates of temperature changes when small changes were made in the watts input to the emitter. In order to obtain this information, a separate electron gun unit containing electrodes A, B, C was made up; the temperature variations were studied using an optical pyrometer focused on the central region of the unipotential cathode. With practice, this could be read to within 3 to 5 centigrade degrees in 1000 centigrade degrees.

It is as well to keep in mind the geometry of the gun when following through and accounting for the current changes observed, particularly in the marginal portions of the beam. The cathode of 1·7 mm diameter when operating is approximately 0·7 mm from the A electrode which has a 1·5 mm aperture. Under these conditions, the electrode would be expected to trim the peripheral region of the beam and to receive a diminishing current with reduction in temperature. Under space charge limited conditions, the fraction of beam current reaching A should remain constant independent of the accelerating voltage.

The question of the effects of secondary electrons released from the electrodes bombarded by the primary beam is important, since even at the low energies used and with previously outgassed metal, the secondary emission coefficient $\delta(\ll 1)$ may affect the electron distribution in the marginal part of the beam, though hardly, if at all, in the core. It is known that δ varies smoothly with primary energy in this experimental range and that the majority of secondary electrons leave the surface with an energy of 2–3 volts. This would lead one to expect that over a small interval of, say, 10 volts secondary emission effects should, if present, produce a smooth variation in the collected current; this is not found to be the case.

It is known that, in order to suppress the secondaries, it is necessary for the space potential outside the secondary emitting surface to be at least three volts negative with respect to the surface. Such a lowering of potential occurs naturally in the present case, for example from the form of the static field in the aperture of the A electrode. Field plot measurements using present potential values indicate that owing to the curvature of the equipotentials in this region (concave to the cathode) there is a reverse radial field of some 30 v mm^{-1} which would in any case prevent secondaries leaving the A electrode and thus from joining the main electron stream. In addition, there is the radial gradient introduced by the space charge of the beam itself. Rough calculation shows that for a beam current of the order of 500 μA the potential at the edge of the beam is reduced by about one volt due to the charge in the core; it is unlikely therefore that any secondaries reach subsequent electrodes. In the field-free space beyond the C electrode, the potential depression produced by the negative charge of the beam at these low energies provides the reverse field inhibiting the secondary emission process. A rough calculation again shows that for a 50 v beam this works out at about 10 v negative; for the lower energies the reverse inhibiting field would be considerably stronger. A check that secondary electrons were absent from the core of the beam was furnished by raising and lowering the potentials of the end collectors individually, through 5 v. This made no change in the current collected by each.

§ 3. EFFECT OF TEMPERATURE: GUN EFFICIENCY

In our earlier paper (1953) it was experimentally established from examination of the beam distributions that under normal emitting temperatures the three-halves power law breaks down below a limiting beam energy. This was manifested by a progressive decrease in axial current until a stage was reached at which it vanished; under these conditions the beam became 'hollow', that is, the current density in the vicinity of the axis became zero, while it remained finite elsewhere. This effect must arise from the formation close to the cathode of a 'virtual' cathode; this must have a peculiar potential distribution across it, compounded of the static distribution on which is superimposed the variable charge density of the beam itself. The potential depression thus formed on the axis is exceedingly sensitive to changes in the energy of the beam, and thus to changes in charge density along and transverse to the beam.

For any fixed modulation ratio such changes may also be brought about by a variation in cathode temperature, since, with the flat topped cylindrical cathode shell used, the emitting area contracts as the temperature is lowered, and differential saturation as between different sections of the beam may occur. The peripheral regions of the emitter which were previously contributing substantially to the total beam current now become more or less inactive as emitting elements, the total current decreases, and, at a sufficiently reduced temperature, saturation sets in and the current becomes independent of the beam energy. However, space charge forces are still operative and may have a marked effect close to the cathode surface where the beam cross section may be greatly dilated. These, and the reduced initial velocity effects, are reflected in the beam distributions as revealed by the series of collectors. The variation of gun efficiency with beam energy for various emitter temperatures is shown in figure 2 for $\alpha = 4$. Reduction of temperature delays the onset of current at the lower energies, due presumably to the presence of an increased barrier effect in the region of the virtual cathode arising from initial thermal velocity variation. This, in turn, requires an increase in the external field for its neutralization, and results in a reduced efficiency owing to the 'splaying out' of the beam; at the higher energies this position is reversed, the smaller space charge effects at the lower temperatures yielding smaller peripheral currents reaching electrodes A, B, C.

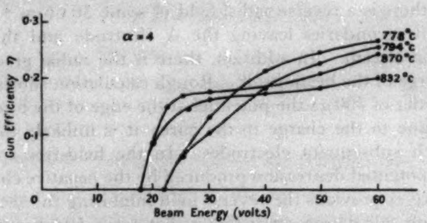


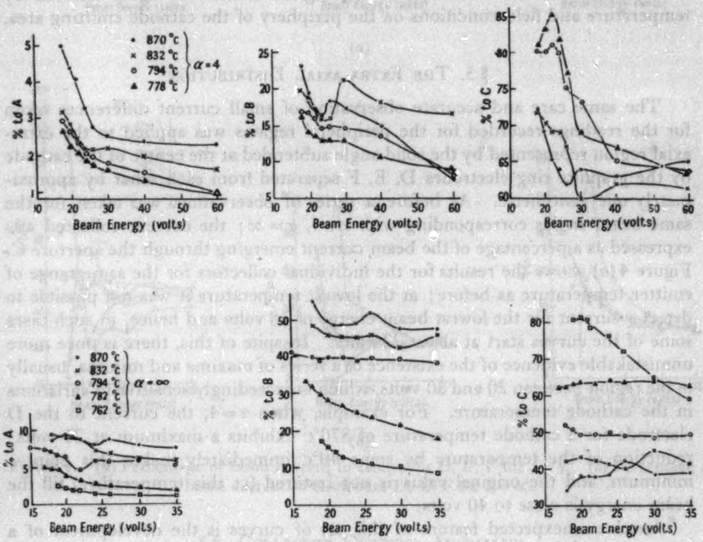
Figure 2. Gun efficiency at various emitter temperatures; $\alpha = 4$.

For the purpose of analysing the beam distribution in some detail, it is proposed to divide it into three portions: (i) marginal, as revealed by the currents to A, B, C electrodes, (ii) extra-axial, as affecting currents to D, E, F electrodes, (iii) axial or core, as collected by the series of concentric collectors 1, 2, 3, 4, 5.

§ 4. THE PERIPHERAL DISTRIBUTION

In the earlier series of experiments recorded in the previous paper (1953) it was found that pronounced maxima and minima had developed in the fractional currents to the trimming electrodes C and B respectively as the accelerating voltage was varied, for particular values of α. In the case of both these electrodes this occurred at the lowest voltage (30 v) and indicated a rather unexpected form for the marginal distribution in that portion of the beam nearest the emitter. This fact led us to make a detailed and accurate exploration of the currents collected by these electrodes for lower beam energies at various temperatures. To avoid missing signs of incipient maxima and minima, the voltage was increased in some cases in steps of 1 or 2 volts at a time, particularly in the region between 17 and 30 volts where it was expected the changes would be most marked. The results are shown in full in figure 3 (*a*) for the case of $\alpha = 4$ and in figure 3 (*b*) for $\alpha = \infty$. In both instances the percentage of the total cathode current is plotted over a temperature variation at the centre of the cathode of about 100°c.

Generally, in contrast to the effects at higher energies where the fractional currents to B and C tend to show a small gradual and smooth fall with increase in energy over the whole energy range (as would be expected if the space charge forces were decreasing) we now get definite maxima and minima in the energy region between 18 and 30 volts for both the B and C electrodes; this is particularly evident for the narrower beam angle, though their presence and position appear to depend on the emitting temperature. If we look at the curves for each electrode in turn, then, from figure 3 (*a*) for the case of the A electrode, there is a steep initial fall; beyond 30 volts the decline is much more gradual, and the rate of fall from this point to the limits of energy is approximately constant at all temperatures. In the interval between 18 and 26 volts at the highest temperature, there is a 50% variation; this in itself indicates in a clear way the remarkable sensitivity of the peripheral distribution to small energy changes in this interval.



(*b*)

Figure 3. (*a*) Percentage of current to A, B, C for $\alpha=4$. (*b*) Percentage of cathode current to electrodes A, B, C for $\alpha=\infty$.

The consistent minimum developed at about 24 volts with the B electrode is maintained as the temperature is reduced till, at the lowest value of 778°c, it has become very broad indeed. Again, the maximum shown between 25 and 30 volts for the two highest temperatures, is seen to be on the point of disappearing at 794°c; at the lowest temperature of 778°c it has not developed at all. The fact that B receives more current at a lower temperature (832°c) above about 27 volts is also interesting as indicative of some peculiar emission pattern which is greatly affected by a combination of field and temperature.

The results with the C electrode are perhaps most interesting of all, and bear out the complicated reaction of field and charge density. Here, it is possible to have large changes of slope in an interval between 18 and 26 volts. At temperatures of 778°c and 794°c, definite maxima have developed; with increase of temperature to 870°c these give way to a pronounced minimum, a rather unexpected and striking result which definitely excludes an explanation based on the simple concept of space charge forces alone.

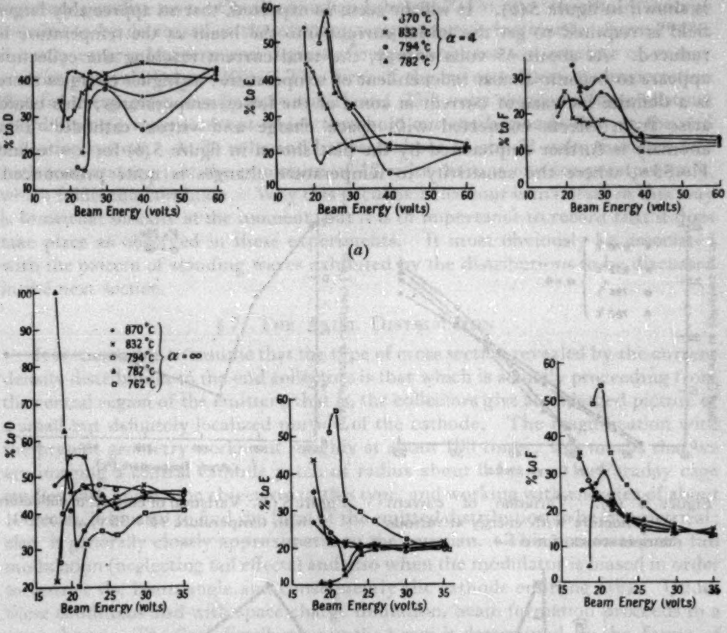
The behaviour as illustrated by figure 3 (*b*), where the initial field conditions are very different ($\alpha = \infty$), conforms to a more orthodox pattern and is not so characteristic as the previous results. As before, both maxima and minima are found to arise in the same interval of 18–26 volts; these, however, are much shallower and are most in evidence at the highest temperatures. It is of some interest to note that these occur even for the A electrode; this, in itself, points to the existence of non-uniform emitting regions which are again affected by temperature and field conditions on the periphery of the cathode emitting area.

§ 5. The Extra-axial Distribution

The same care and accurate observation of small current differences taken for the readings recorded for the peripheral regions was applied to the extra-axial region represented by the solid angle subtended at the centre of the cathode by the graphite ring electrodes D, E, F separated from each other by approximately one centimetre. As before, a series of observations was taken for the same beam angles corresponding with $\alpha = 4$, $\alpha = \infty$; the current collected was expressed as a percentage of the beam current emerging through the aperture C. Figure 4 (*a*) shows the results for the individual collectors for the same range of emitter temperature as before; at the lowest temperature it was not possible to detect a current for the lowest beam energy of 18 volts and hence, in such cases some of the curves start at about 22 volts. In spite of this, there is once more unmistakable evidence of the existence of a series of maxima and minima, usually in the region between 20 and 30 volts, which is exceedingly sensitive to variations in the cathode temperature. For example, when $\alpha = 4$, the current to the D electrode for a cathode temperature of 870°c exhibits a maximum at 21 volts; reduction of the temperature by some 40°c immediately makes this point a minimum, and the original value is not restored (at this temperature) till the beam energy is close to 40 volts.

Another unexpected feature of this set of curves is the development of a minimum at about 40 volts which apparently only occurs at temperatures of 870°c and 794°c and not at the intermediate temperature of 832°c. The exceptionally steep initial rise at the lower temperatures within an interval of 2 volts is rather surprising; it serves to emphasize the sensitivity accompanying field variations over small intervals at the 'threshold'. All points tend to come together at 60 volts independent of temperature; this would appear to show that a beam of this energy is beginning to behave in normal fashion.

These general characteristics are also found for electrodes E and F, though the former does not show such prominent variations (the absence of readings at the lowest beam energies at the lower temperatures might account for this); the maxima and minima are better developed for the F electrode. For the larger beam angle involving greater current densities, figure 4 (*b*) shows that as compared with the previous set of curves, most of the changes occur within a narrower energy interval, 18–25 volts. The same abrupt rise and fall takes place over a small energy range, and, judging from the curves for electrodes E and F, the beam begins to 'behave' in these regions at lower final energies of about 30 volts.



(*a*)

(*b*)

Figure 4. (*a*) Percentage of beam current to electrodes D, E, F for $\alpha = 4$. (*b*) Percentage of beam current to electrodes D, E, F for $\alpha = \infty$.

§ 6. COLLECTOR CURRENT ANOMALIES

While the interesting variations so far discussed for the extra-axial regions have revealed unexpected periodic fluctuations over small energy intervals round critical energies of 20 volts, most interest was however centred on the distribution in the core of the beam. The system of almost coplanar collecting discs of gradually increasing diameter served to resolve this sufficiently closely, to show up fine differences in current density along a radius, starting from the axis of the system and proceeding outwards towards the periphery of the collectors. It was also possible to link up with one or more points of the extra-axial distribution obtained for the electrodes D, E, F by projecting their areas on to the plane of the end collectors; this revealed how the form of the current density distribution developed at large distances from the axis.

Before commenting on these curves in the next section, it is as well to present some features of the total current reaching the collectors as a unit, under various conditions of energy and temperature, since it exhibits certain anomalies which require to be kept in mind when interpreting the distribution curves. A 'characteristic' current–voltage curve for $\alpha = 4$, at various cathode temperatures is shown in figure 5 (*a*). It will be seen, as expected, that an appreciably larger field is required to get the initial current into the beam as the temperature is reduced. At about 45 volts energy, the total current reaching the collectors appears to become almost independent of temperature; at higher energies there is a definite increase of current at some of the lower temperatures; this could arise from effects connected with space charge and virtual cathode. This anomaly is further emphasized by the plot shown in figure 5 (*b*) for $\alpha = \infty$ and $V = 35$ v, where the sensitivity to temperature changes is quite pronounced. In this case a considerably larger total current reaches the collectors at the lowest temperature and a minimum appears to have developed at about 800°c; no variation occurs between 830 and 870°c. Coupled with this effect of temperature, it was of considerable interest to examine the combined effect of field and temperature (particularly in the region of beam energies between 18 and 24 volts,

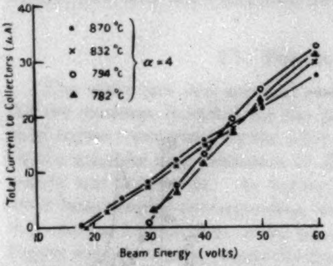


Figure 5 (*a*). Variation of current to collectors with energy at various temperatures, for $\alpha = 4$.

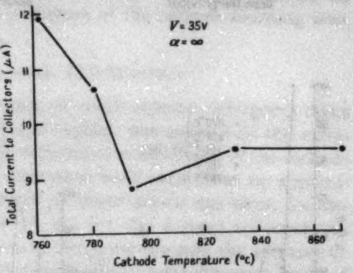


Figure 5 (*b*). Variation of current to collectors with temperature ($V = 35$ v, $\alpha = \infty$).

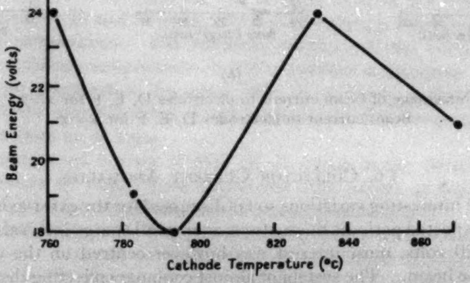


Figure 5 (*c*). Threshold current to collectors at different temperatures for various beam energies, for $\alpha = \infty$.

at which marked development of maxima and minima had occurred in the earlier curves for the extra-axial distributions) in producing a 'threshold current' to the collectors. It was hardly possible to predict the variations which actually occurred as shown in figure 5 (*c*), for $\alpha = \infty$. The outstanding result of this experiment is to indicate that both field and emission density are inter-related in a very peculiar way. On the one hand, below 794°c, the 'threshold' energy requires to be decreased as the emitting temperature is increased; above 794°c, the very reverse happens and there are definite minima and maxima of energy over an interval of some 30°c. The form of the curve leads to the expectation that for every given energy between 21 and 23·8 volts there are three distinct and differing temperatures at which threshold current begins to flow to the end collecting system. At 24 volts and between about 18 and 21 volts there are two such temperatures; at the minimum of 18 volts there is one temperature only, which fulfils this condition. Why this peculiar behaviour is manifest in this way is somewhat obscure at the moment, but it is of importance to record that it does take place as observed in these experiments. It must obviously be associated with the pattern of standing waves exhibited by the distributions to be discussed in the next section.

§ 7. THE AXIAL DISTRIBUTION

It is reasonable to assume that the type of cross section revealed by the current density distribution to the end collectors is that which is actually proceeding from the central region of the emitter; that is, the collectors give an enlarged picture of a small but definitely localized portion of the cathode. The magnification with the present geometry works out roughly at about 100 times; this means that we are imaging a central cathode patch of radius about 0·1 mm. In Faraday cage studies with immersion objectives of this type, and working with energies of about 100 volts, it is usual to find the form of the emitted distribution to be symmetrical; also, it generally closely approximates to the gaussian. This occurs at both full modulation (neglecting tail effects) and also when the modulator is biased in order to restrict the beam angle and consequently the cathode emitting area. Under these conditions and with space charge limitation, beam formation proceeds in a normal way. The final distribution in the beam is determined by the forms of (*a*) the electrical field, (*b*) the velocity distribution, (*c*) the space charge forces. It is of course true to say that there will be inherent variations of emission from point to point of the emitter (for the same field conditions at each) depending among other things on the form and coherence of the oxide coating; however, exploration of this property using an emission microscope shows that, providing the vacuum conditions do not deteriorate, and that the cathode is worked below a safe loading, these characteristic variations occur over the whole surface and can be repeated at will. This indicates that once the surface is brought into a certain condition, it should not change appreciably over long periods of time.

With the very low energy beams with which we are dealing, we should expect to find initial velocity effects quite marked. These produce a potential minimum close to the cathode surface, and provide a potential barrier of sufficient magnitude and depth to restrict the emission over the central area of the cathode in a way determined by its form and extent. Rough calculation from the ratio of beam to saturation currents shows the potential minimum to be between the limits $-0{\cdot}5$ and $-0{\cdot}4$ volt for final energies of 20–30 volts respectively; and, when it is remembered that the most probable emission velocity is about 0·1 volt at the highest temperature used, there can be no doubt about the effectiveness of the reflecting barrier. A further rough calculation puts the position of this potential minimum at about 0·01 mm from the surface, so that this represents an approximate value for the limiting range of the electrons emitted from a small area confined mainly to the centre of the cathode, since the marginal densities are small. The axial distribution curves lend support to these views because they indicate the existence of semi-hollow beams, that is, those for which the current density is larger at points off the axis than at points on the axis. The form and character of such distributions for close differences in energy and temperature are shown in figure 6 (*a*) and (*b*) for $\alpha = 4$ and ∞ respectively. The point of interest in figure 6 (*a*) is the appearance of maxima of the current density off the axis for energies of 20 and 22 volts. These are maintained for the 20-volt beams throughout the temperature range studied, though the curve for 870°c shows the existence of a second maximum further out from the axis. At and above 24 volts energy, the

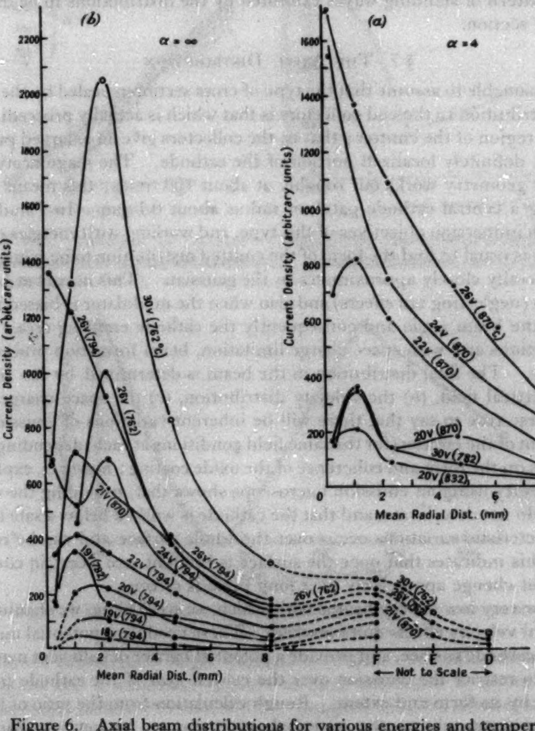


Figure 6. Axial beam distributions for various energies and temperatures, (*a*) $\alpha = 4$, (*b*) $\alpha = \infty$.

distribution becomes 'normal' for the highest temperature; if, however, the temperature is reduced from 870°c to 782°c it is possible to produce these semi-hollow beams at the higher energies, as shown by the 30-volt curve.

These main conclusions are confirmed in figure 6 (*b*) where greater current densities and lower beam energies were investigated. The interest here lies in the demonstration that it is possible to change a maximum of current into a minimum, and also introduce a second maximum simply by introducing a 2-volt variation in energy at a temperature of 794°c; this appears to occur at a critical energy of about 24 volts. Below this value, the 2-volt energy variation does not affect the form of the distribution, it merely increases the current density at the maximum already formed at lower energies; above this, the periodic character of the variation disappears completely and the curve falls away smoothly as for a normal distribution. As previously stated, it is possible to reintroduce the maximum at the higher energies by a decrease of temperature by some 30°c as shown by the curves for 26 volts (and also 30 volts) at the temperature of 762°c; here, the maximum is much more marked, but is, however, displaced further from the axis.

If the current density distribution is followed further from the axis by projecting the areas of the F, E, D electrodes on to the plane of the collectors, a second maximum appears in process of development at the F electrode as shown in figure 6 (*b*)—the points marked F, E, D along the axis are not to scale on this curve—for all the lower energies at all temperatures. Inclusion of a number of other results not shown in these diagrams leads, for $\alpha = \infty$, to a classification such as the following :

Temperature (°c)	Normal Type Distribution Beam Energy (v)	Periodic Type Distribution Beam Energy (v)
870	22, 24, 26, 30, 35	21
832	24, 26, 30, 35	
794	26, 30, 35	19, 20, 22, 24
782	20, 22, 24, 26, 30, 35	19
762	35	24, 26, 30

The general conclusions to be drawn from this analysis are :

(i) reduction in temperature may result in introducing abnormalities into the higher energy distributions;

(ii) the change from normal to abnormal may occur over a 1-volt energy interval;

(iii) an energy change of 2 volts can alter the periodic pattern and convert a minimum into a maximum;

(iv) two maxima may occur at a critical energy; reduction of temperature causes one to disappear;

(v) where a maximum is recorded for the collectors, a second maximum usually appears further out from the axis.

This periodic variation in emission across the cathode surface can be interpreted as arising from the effect of standing waves set up in the potential well immediately in front of the cathode. This is brought about by the effects of initial electron velocity as well as the space charge. It results in the conversion of the type of field common with immersion objectives; that is, one in which the gradient of potential is positive outwards from the virtual cathode, to one characteristic of a saddle-back field in which the gradient is oppositely directed on each side of the saddle point. The electron oscillations take place at right angles to the axis; they react on the emission distribution to produce the irregularities recorded by the experiments. This type of behaviour conforms with the observations of Klemperer (1947) on the splitting of high energy electron pencils into 'elementaries' giving a standing wave pattern in the space charge near the emitter. The presence of a deep potential trough surrounding the saddle point, which must occur for the very low energy beams considered in this work, gives a mechanism for the production of oscillations in the deep well. A good analogy exists in the case of the transition of the symmetrical electron lens to 'symmetrical' mirror; here the change-over is exceedingly sensitive to small variations in potential at the saddle point. The electron, in place of passing through the lens field as a focused ray, now begins to oscillate at right angles to the axis in the region surrounding the saddle point. This type of motion may be changed to direct reflection (or transmission) by slight alteration of the field conditions.

As the minimum wavelengths involved are determined by consideration of the patterns of minima and maxima and are of the order of 0·01 mm, one would expect the potential barrier to be found at approximately this distance from the emitter. This is indeed found to be the case from a rough calculation of its position for the equivalent diode.

The distance between cathode and potential minimum would then be of the order of half the wavelength and it is also possible for overtones of the fundamental frequency to arise, which if superimposed on the fundamental standing wave pattern would make the latter exceedingly sensitive to temperature changes. That this mechanism is essentially correct can be seen by considering an alternative explanation based on electron oscillation in an electron plasma cloud in front of the cathode. Taking rough values for the emission density as 0·03 A cm^{-2} and a most probable velocity of 0·1 ev the charge density works out at about $1{\cdot}5 \times 10^{-9}$ coulomb cm^{-3}. This corresponds to an electron density of 10^{10} electrons per cm^{3} and a plasma frequency of 9×10^{8} c/s having an equivalent wavelength of some 30 cm. This value is some four orders of magnitude too large and a mechanism of this type must be discarded. Further work hence remains to be done to get more quantitative results for the form, character and detailed effects arising from the saddle-back field.

ACKNOWLEDGMENT

In conclusion the authors wish to thank Professor H. W. B. Skinner for his interest in this work.

REFERENCES

CLARKE, W. H. H., and JACOB, L., 1951, *Nature, Lond.*, **168**, 1120; 1953, *Proc. Phys. Soc.* B, **66**, 284.

KLEMPERER, O., 1947, *Proc. Roy. Soc.* A, **190**, 376.

A Possible Chromatic Correction System for Electron Lenses

By G. D. ARCHARD
Electron Physics Section, Associated Electrical Industries, Ltd., Aldermaston, Berks.

MS. received 25*th February* 1955, *and in amended form* 12*th April* 1955

Abstract. It is shown that a suitably arranged group of four electrodes has properties which make it a useful integral part of a system for the correction of axial chromatic aberration. Its application to the correction system proposed by Scherzer, and to a new simplified system having certain advantages over that of Scherzer is discussed. Curves are given for the chromatic and focal properties of the corrective electrode group, and an example of a corrective system is considered numerically. The effect of this system on off-axial chromatic aberration is investigated, and the possibility of combining it with a system for spherical correction is envisaged.

§ 1. INTRODUCTION

It is well known that non-rotationally symmetric arrays of electrodes have properties which may be applied to the correction of certain aberrations of electron lenses. Thus, an array of four equal spheres bearing alternate equal and opposite charges, distributed at intervals of 90° round the axis of an electron beam, with their centres equidistant from the axis and lying in a plane perpendicular thereto, may be used to correct first-order astigmatism due to the mechanical defects of a preceding electron lens (Rang 1949). Again, several groups of eight equal alternately charged spheres, the spheres in each group being distributed at intervals of 45° round the axis of the beam with their centres equidistant from the axis and lying in a plane perpendicular thereto, may be used in conjunction with certain cylindrical lenses to correct third-order spherical aberration (Scherzer 1947, Seeliger 1951). In either case, the spheres may be replaced by electrodes of other shapes.

It will be shown in the following that four equal alternately charged spheres distributed at intervals of 90° round the axis of the beam, having their centres in one plane perpendicular to the axis but *not* necessarily being equidistant therefrom, possess properties which may be applied to the correction of *chromatic aberration*. Such an array will be called a ‘four-pole group’. In the particular case where two opposite spheres recede to infinity and the others remain at equal finite distances from the axis, this will become a ‘two-pole group’.

The importance of the correction of chromatic aberration lies in the empirical fact that it is a major feature in limiting the attainment of high resolution in the electron microscope. This has not always been regarded as true. In 1943 Gabor stated (Gabor 1944) that chromatic and spherical aberrations might even be *beneficial* to the production of good images in transmission microscopy. However, this statement was made at a time when electron microscopy was in its infancy, and it was not accompanied by any experimental evidence that a particular microscope would be improved when chromatic or spherical aberration were deliberately introduced. Indirect evidence *is* now available, for example Boersch (1953) found considerable improvement in electron micrographs when a velocity filter was employed to remove inelastically scattered electrons. Boersch pointed out that, whereas his experiments carried out with gold foil showed appreciable improvement, an even greater improvement would be expected with biological specimens. Leonhard (1954), who has still more recently made an extensive experimental study of electron scattering has found that electron diffraction photographs also show greater sharpness and contrast when a filter is used.

Chromatic defects are, however, much more important in the *dark field* microscope (Gabor 1944, 1951), and in the *reflection* microscope (Kushnir, Bibermann and Levkin 1951, Haine and Hirst 1953), in which they are believed to be largely responsible for a fall in resolution of an order of magnitude as compared with transmission microscopy. Reflection microscopy is growing in importance, but in order to compare in resolution with transmission microscopy it is essential that the reflection microscope should include at least some form of chromatic correction. However, it is always possible that future investigators may discover circumstances under which a measure of chromatic aberration is, as originally suggested by Gabor, desirable in *transmission* microscopy. In that case the correction system about to be described might assume an even greater relevance, for its function would then be to introduce a controllable measure of chromatic aberration into the lens system (by varying the voltages on its electrodes), and the optimum measure could then be systematically and quantitatively sought.

Rather less attention has been given to chromatic correction than to the correction of spherical aberration, which, though important, is generally masked by chromatic effects. Scherzer (1947) demonstrated mathematically the possibility of correcting axial chromatic aberration by means of non-rotationally symmetric potential distributions set in a region in which the electron beam had been rendered deliberately astigmatic, but no attempt seems to have been made to suggest a useful practical form for the corrective system proposed.

It will be shown in the following that a combination of three two-pole groups would have nearly the right potential distribution for Scherzer's system, but that a combination of two two-pole groups, though not obeying Scherzer's potential distribution, would have essentially the same corrective effect. It will further be shown that the form and means of production of the astigmatic beam can be simplified, and that, whereas Scherzer's original potential distribution would not suit the modified beam, the potential distribution of two two-pole groups will do so and will lead to the achievement of complete axial chromatic correction. Chromatic correction for points not on the axis will not be achieved by this system, but it will be shown that, for a restricted field, purely axial correction will result in improved imaging.

§ 2. POTENTIAL DISTRIBUTION OF FOUR-POLE AND TWO-POLE GROUPS

Consider the four-pole group of figure 1. The z axis represents the axis of the beam. Two spheres of radius ρ bearing charges $+q$ have their centres on the x axis at distance a from the origin. Two spheres of radius ρ bearing charges $-q$ have their centres on the y axis at distance b from the origin. It may readily be shown that, for values of r much smaller than a, b (as will generally be in question when dealing with electron beams) and for ρ very much less than a, b the potential at a point (r, θ, z) $(x=r\cos\theta, y=r\sin\theta)$ is

$$\phi=\Phi-\tfrac{1}{4}\Phi''r^2+\Phi_2 r^2\cos 2\theta+O(r^4) \qquad \ldots\ldots(1)$$

where

$$\Phi=\Phi_A+2\Phi_S\rho[(a^2+z^2)^{-1/2}-(b^2+z^2)^{-1/2}] \qquad \ldots\ldots(2)$$

$$\Phi_2=\tfrac{3}{2}\Phi_S\rho[a^2(a^2+z^2)^{-5/2}+b^2(b^2+z^2)^{-5/2}]. \qquad \ldots\ldots(3)$$

Φ_A is the accelerating voltage, Φ_S the voltage on the spheres, and dashes indicate differentiation with respect to z. This is in conformity with Scherzer's (1947) notation. The proviso that $\rho \ll a, b$ is made so that the capacities of the spheres may be replaced by their radii, but the relations may be accepted as reasonable approximations provided that $\rho/a, \rho/b \ngtr \tfrac{1}{2}$.

A rotationally symmetric ('round') electron lens has a potential distribution of form $\phi=\Phi-\tfrac{1}{4}\Phi''r^2+O(r^4)$ where Φ is a function of z only. On the other hand, a symmetrical four-pole group $(a=b)$ has

$$\phi=\Phi_2 r^2\cos 2\theta, \qquad \Phi_2=3\Phi_S\rho a^2(a^2+z^2)^{-5/2}.$$

Thus the four-pole group in general represents a combination of a round lens and a symmetrical four-pole group. The properties of the latter have been discussed elsewhere (Archard 1955 b). In the 'weak lens approximation', a symmetrical four-pole group deflects electrons travelling respectively in the xz and yz planes equally and oppositely. A round lens, however, deflects both kinds of electron equally and (except for an irrelevant case) always in a convergent sense. Thus, the general four-pole group may be expected to cause electrons in one of the said planes to diverge but electrons in the other plane to converge more than the former diverge, so that the mean deflection of the two kinds is towards the axis.

In particular, the potential distribution of a two-pole group will be found by putting either a or b equal to ∞ in equations (2) and (3).

§ 3. RAY PATHS IN A TWO-POLE GROUP

Scherzer (1947) gave equations for the gaussian ray paths of electrons in a field whose potential had the form (1) in which both Φ and Φ_2 were arbitrary functions of z. The equation for the ray travelling in the xz plane was

$$\Phi x''+\tfrac{1}{2}\Phi'x'+(\tfrac{1}{4}\Phi''-\Phi_2)x=0. \qquad \ldots\ldots(4)$$

If the values of Φ and Φ_2 given by (2), (3) with $a=\infty$ are substituted in (4), there results

$$[(\Phi_A/\Phi_S)-2\rho(b^2+z^2)^{-1/2}]x''+\rho z(b^2+z^2)^{-3/2}x'-\rho(b^2+z^2)^{-3/2}x=0. \qquad \ldots\ldots(5)$$

If $\xi=b/(b^2+z^2)^{1/2}$, this becomes

$$(d^2x/d\xi^2)(\xi-2v\xi^2-\xi^3+2v\xi^4)+(dx/d\xi)(2-5v\xi-3\xi^2+7v\xi^3)-xv=0 \qquad \ldots\ldots(6)$$

where $v=(\rho/b)(\Phi_S/\Phi_A)$.

This may be solved in series, and the result, up to the second power of v, is

$$x=A[1+v\{(1+z^2/b^2)^{1/2}-z/b\}+v^2\{\tfrac{1}{2}(1+z^2/b^2)^{-1}+\ldots\}] \\ +B[(z/b)-v+v^2(z/b)]. \qquad \ldots\ldots(7)$$

To the order v^2, this constitutes the exact solution for the ray paths in the xz plane. Those in the yz plane will be less material, as will appear. Equation (7) checks with the obvious solution $x=A+B(z/b)$ for $v=0$.

§ 4. FOCAL PROPERTIES OF A TWO-POLE GROUP

For x to be finite at $z=-\infty$ (so that the ray approaches the two-pole group parallel to the axis) (7) requires that $B=2Av$ so that to order v^2:

$$x=A[1+v\{(1+z^2/b^2)^{1/2}+z/b\}+v^2\{\tfrac{1}{2}(1+z^2/b^2)^{-1}+\tfrac{1}{8}(1+z^2/b^2)^{-2}+\ldots\}-2v^2] \qquad \ldots\ldots(8)$$

whence

$$x'=A[v\{(z/b^2)(1+z^2/b^2)^{-1/2}+1/b\}+v^2\{-(z/b^2)(1+z^2/b^2)^{-2}-\ldots\}]+\mathrm{O}(v^3). \qquad \ldots\ldots(9)$$

As $z\rightarrow\infty$, so

$$x'\rightarrow 2Av/b+\mathrm{O}(v^3), \qquad \ldots\ldots(10)$$

all the v^2 terms vanishing. Thus, even to a second order approximation, the change of slope caused by the two-pole group is

$$x'=2Av/b \qquad \ldots\ldots(11)$$

in which A is identified with the height of the incident ray. Plainly, v can have either sign (following Φ_S); hence, the two-pole group may either diverge or converge rays in the xz plane. Comparison with the properties of the symmetrical four-pole group given elsewhere (Archard 1955 b) shows that the two-pole group has, for rays in the xz plane, exactly half the deflecting power of the corresponding symmetrical four-pole group, in the weak lens approximation.

§ 5. CHROMATIC PROPERTIES OF THE TWO-POLE GROUP

Scherzer has shown (1947) that when a ray travelling in the xz plane passes through a field characterized by the quantities which have been designated Φ, Φ_2, there occurs a chromatic image displacement of

$$x_1=\frac{M\alpha\epsilon}{\sqrt{\Phi_A}}\int_{z_0}^{z_1}\left(\frac{\Phi_2}{\Phi}-\frac{3}{8}\frac{\Phi'^2}{\Phi^2}\right)\frac{x_\alpha^2}{\sqrt{\Phi}}\,dz \qquad \ldots\ldots(12)$$

where M = magnification, α = angular semi-aperture, ϵ = small perturbation of Φ representing chromatic kinetic energy, x_α is a ray starting from a point z_0 on the axis at slope unity, and z_1 represents the image plane. The chromatic properties of the two-pole group will be considered on the assumption that it is to follow a round lens (e.g. an electron microscope objective lens) so that x_α may be regarded as entering the region of influence of the two-pole group at height A (approximately equal to the focal length of the round lens) and zero inclination.

If the values of Φ, Φ', Φ_2 obtained from equations (2) and (3) with $a=\infty$ be inserted in (12) and the integral be evaluated, there results

$$x_1=\frac{M\alpha\epsilon}{\Phi_A}\frac{A^2}{b}\frac{3}{2}v\left\{\frac{4}{3}+v\left(2\pi+\frac{4}{3}\right)\right\}$$

or, in terms of the effective chromatic aberration coefficient C_c,

$$C_c=\frac{A^2v}{b}\{2+v(2+3\pi)\}. \qquad \ldots\ldots(13)$$

Unlike x', C_c contains an appreciable term in v^2, i.e. it departs considerably from a linear dependence on v; when positive, it tends to correct (in the xz plane) the chromatic aberration of the preceding round lens.

If now a combination of two two-pole groups be formed (figure 1 (b)), the two component groups having the same value of b but opposite values of v, say $\pm v^*$, and if they are not close enough appreciably to modify the properties which have been deduced for each separately, the changes of slope produced by the respective groups will cancel, but there will be left a combined chromatic coefficient

$$C_c = 2(A^2 v^{*2}/b)(2+3\pi) \qquad \ldots\ldots(14)$$

which is essentially positive and so has the right sense for correction. This is in some ways analogous to the optical 'direct-vision spectroscope', save that, in this case, the two two-pole groups have been combined not to produce dispersion but to annul that which already exists. As in the case of optics, the system could be made more symmetrical by splitting one of the groups into two, each having half the value of v, so forming a combination of three two-pole groups.

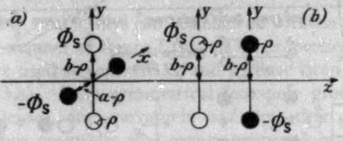


Figure 1. (*a*) Four-pole group, (*b*) combination of two-pole groups.

A combination of two (or three) two-pole groups of this kind may be used to produce chromatic correction (in one plane) together with either convergence or divergence, simply by varying the relative values of $v(=(\rho/b)(\Phi_S/\Phi_A))$. By itself, however, its value would be limited to such devices as require chromatic correction in only one plane. For the complete chromatic correction of a round lens, the existence of an *astigmatic* beam and the insertion of 'corrective lenses' therein are essential, as has been shown by Scherzer (1947). This will be considered in the next section. Meanwhile the approximations which have been made will be summarized: (i) higher order geometric and chromatic aberrations of the correcting devices have been assumed to be negligible, as they are in uncorrected lenses; (ii) relativistic effects have been assumed negligible; (iii) it has been assumed that v^3 may be neglected in comparison with v, and that $A(1-2v^2)$ may be replaced by A where this merely signifies the height of entry of a ray into a correcting device.

§ 6. SCHERZER'S SYSTEM FOR CHROMATIC CORRECTION

In this section, Scherzer's (1947) requirements for chromatic correction in all planes will be reviewed.

Corresponding to (4), Scherzer showed that rays travelling in the yz plane would obey

$$\Phi y'' + \tfrac{1}{2}\Phi' y' + (\tfrac{1}{4}\Phi'' + \Phi_2)y = 0 \qquad \ldots\ldots(15)$$

and, corresponding to (12), there would be an image displacement of

$$y_1 = \frac{M'\alpha\epsilon}{\sqrt{\Phi_A}} \int_{z_0}^{z_1'} \left(-\frac{\Phi_2}{\Phi} - \frac{3}{8}\frac{\Phi'^2}{\Phi^2} \right) \frac{y_\alpha^2}{\sqrt{\Phi}}\, dz \qquad \ldots\ldots(16)$$

y_α being a ray starting from a point z_0 on the axis at slope unity. M' and z_1' are the magnification and image position corresponding to the yz plane. Even when distortion is made zero by setting $M = M'$ and astigmatism is made zero by setting $z_1 = z_1'$, the right-hand sides of equations (12) and (16) cannot both be made to vanish simultaneously if either $\Phi_2 = 0$ or $x_\alpha^2 = y_\alpha^2$ throughout the range of integration. Scherzer therefore proposed to deflect the rays x_α and y_α differently on their emergence from the lens under correction (L_1, figure 2 (*a*)) by means of a pair of convergent cylindrical ('slit') lenses, so as to make them cross the axis at different values of z; at the point where each crossed the axis, he proposed to insert some form of corrective lens which would introduce a corrective chromatic displacement into the other ray, but would leave the directions of both rays unchanged. He showed that such a lens could be formed, for example, if an electrode system could be constructed having

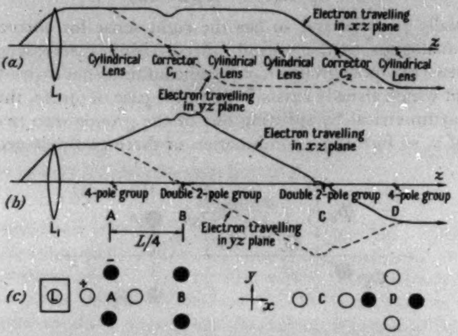

(a)
L₁
Cylindrical Lens
Corrector C₁
Cylindrical Lens
Cylindrical Lens
Corrector C₂
Cylindrical Lens
Electron travelling in xz plane
Electron travelling in yz plane
z
(b)
L₁
4-pole group
Double 2-pole group
Double 2-pole group
4-pole group
A
B
C
D
L/4
Electron travelling in xz plane
Electron travelling in yz plane
z
(c)
L
+
−
A
B
C
D
y
x

Figure 2. (*a*) Scherzer's proposed system, (*b*) simplified system, (*c*) electron's view in various planes of (*b*).

$$\Phi_2/\Phi = (\Phi''/2\Phi) - (\Phi'/4\Phi)^2. \qquad \ldots\ldots(17)$$

This would give an exact solution for (4) in the form

$$x = \Phi^{1/4}[C + D \int \Phi^{-1}\,dz]. \qquad \ldots\ldots(18)$$

If the sign of Φ_2 in (17) were changed, a similar solution would be provided for (15).

Thus, in the arrangement of figure 2 (*a*), x_α would emerge from the lens under correction at height f (focal length of lens) and slope zero; it would be undeflected by the first cylindrical lens, but on entering the first corrective lens C_1 would follow a path

$$x_\alpha = f(\Phi/\Phi_A)^{1/4} \qquad \ldots\ldots(19)$$

emerging again at height f and zero inclination, having introduced a chromatic image displacement, in the corrective sense, of

$$x_1 = \frac{M\alpha\epsilon}{\sqrt{\Phi_A}} \int \frac{1}{16}\left(\frac{\Phi'^2}{\Phi^2}\right)\frac{x_\alpha^2}{\sqrt{\Phi}}\,dz, \qquad \ldots\ldots(20)$$

integration being taken over the corrective lens only.

Meanwhile, y_α would have crossed the axis under the influence of the first cylindrical lens, been unaffected by C_1 (in the region of which it is zero or very small—this will be checked later, §9) and been made parallel to the axis again by the second cylindrical lens. A like system, involving C_2, would then give a corrective displacement to y_α, at the same time reuniting x_α with it—the latter function being necessary in order to avoid first-order distortion. (The word 'reuniting' must be interpreted with caution, as x_α travels in the xz plane and y_α in the yz plane). The complete system would thus counteract the chromatic aberration of the lens under correction (given by (16) with $\Phi_2=0$ and integration extending only over the lens).

It is now plain that *combinations of two-pole groups, such as that discussed in the preceding section, could be used to replace* C_1 *and* C_2, for, although the pair of two-pole groups does not obey (17), it can provide the same essential properties, that is, chromatic correction in one plane (cf. (20)) without change of inclination (cf. (19)).

However, a further simplification is envisaged in the following section.

§ 7. Simplified Ray Separation

Separation of the x_α and y_α rays can be achieved without the use of cylindrical einzel lenses, in the manner of figure 2 (*b*). The properties of the symmetrical four-pole group here used to separate x_α and y_α have been discussed elsewhere (Archard 1954, 1955 b). Each symmetrical four-pole group acts as a pair of crossed cylindrical lenses, one convergent and the other divergent. The ray separation system thus resembles that earlier proposed in connection with the correction of spherical aberration (Archard 1955 a).

One advantage of this modification is that a given $x_\alpha - y_\alpha$ separation can be achieved with a much lower voltage than that required for a cylindrical einzel lens (see for example Archard 1955 a, b). Another advantage lies in the fact that the ray x_α reaches an off-axis distance of $2f$ at the same time as the ray y_α reaches the axis, as opposed to a distance of f in the system of Scherzer, thus increasing the chromatically corrective effect by a factor of four. (The two deflecting symmetrical four-pole groups have equal and opposite values of Φ_S, and so cancel each other's contributions to the integrals (12) and (16).)

It is clear in the system depicted, however, that the corrective 'lenses' must not only insert a chromatic correction, but must cause the electrons to converge as well. This could not have been done with Scherzer's corrective lenses as characterized by (17), as these essentially fail to change the *slope* of the rays. It may be shown by a perturbation method, however, that, if a little extra Φ_2 is added at the centre of each of Scherzer's corrective lenses, and a little Φ_2 is removed from their flanks, the modified corrective lens will provide convergence without losing its chromatically corrective properties. This could be achieved, for example, by the use of the three two-pole groups depicted in figure 3 (*a*), for, as is seen from figure 3 (*b*), their values of $\Phi''/2$ and Φ_2 practically coincide so that (17) is nearly obeyed (the term in Φ' being small); slight variations in their potentials could therefore effect the desired adjustments in the distribution of Φ_2. Nevertheless, this procedure would be unnecessary, as it has already been seen that convergence and correction can be achieved simultaneously with a combination of only two two-pole groups (§ 5).

With $A=2f$ in (10) and (13) therefore, there result the following expressions for the change of slope $x_\alpha{}'$ and chromatic correction coefficient C_c due to a single two-pole group:

$$s=\frac{x_\alpha{}'}{f/b}=4v+\mathrm{O}(v^3) \qquad \ldots\ldots(21)$$

$$c=\frac{C_c/f}{f/b}=8v+4(2+3\pi)v^2+\mathrm{O}(v^3); \qquad \ldots\ldots(22)$$

these will be used for the numerical calculation of the properties of double two-pole groups.

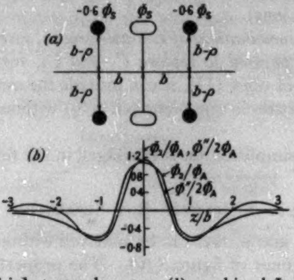

-0·6 ϕ_s
ϕ_s
-0·6 ϕ_s
(a)
$b-\rho$
$b-\rho$
b
b
$b-\rho$
$b-\rho$
(b)
1·2
$\phi_s/\phi_A, \phi''/2\phi_A$
0·8
ϕ_s/ϕ_A
0·4
$\phi''/2\phi_A$
-3
-2
-1
1
2
3
z/b
-0·4
-0·8

Figure 3. (*a*) Triple two-pole group, (*b*) combined Φ_s/Φ_A and $\Phi''/2\Phi_A$.

§ 8. Numerical Results

The *s* and *c* values of a single two-pole group are given by equations (21) and (22). Two such groups with equal and opposite values of *v* yield the curves for *c* and *s* marked A in figure 4. That for *s* is merely the line $s=0$, and this would only be of value in Scherzer's unmodified system (figure 2 (*a*)) which requires no change of slope on the part of the corrective group. The remaining curves, B to E, for *c* and *s* correspond to other relative values of the *v*'s of the two two-pole groups; in these cases, *s* is always negative (yielding convergence) so the curves apply to the modified system of figure 2 (*b*).

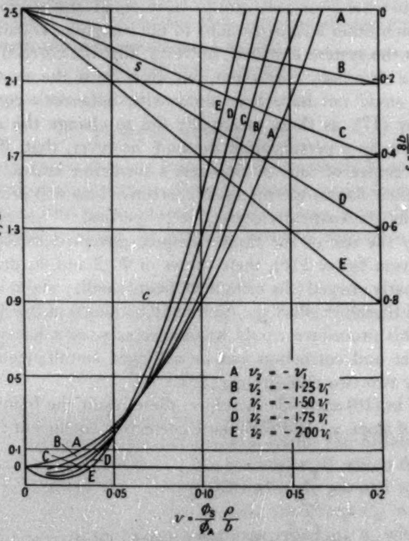


Figure 4. Chromatic and focal properties of a double two-pole group.

The total length of the astigmatic beam, which forms the greater part of the length of the correction system, is connected with s by the formula $L=8b/s$ and the chromatic correction as before is given by $C_c=f^2c/b$.

To take a typical case, if $f=0\cdot5$ cm, $b=0\cdot25$ cm, $\rho=0\cdot125$ cm, $v_2=-0\cdot15$, $v_1=0\cdot10$, the length L will be 10 cm, and the chromatic correction $C_c=1\cdot1$ cm, while $\Phi_{S1}=0\cdot1\Phi_A$ and $\Phi_{S2}=-0\cdot15\Phi_A$. The spheres of radius 0·125 cm may be replaced by laminae 0·5 cm apart (in a direction perpendicular to the beam) extending 0·8 cm parallel to the beam (Archard 1955 b). Care must be taken, however, in such replacements, to see that there is enough room left to accommodate the laminae. If b and both Φ_S's are halved, the v's remain the same, so that C_c and $1/L$ are doubled. L and C_c may be restored to their values by lowering the v's. It is thus advantageous from the point of view of voltages to reduce the gap between the corrector electrodes, provided that it is borne in mind that tolerances vary as b^4 (Archard 1954, 1955 *a*). The dimensions given above correspond to tolerances of the order of 20 to 30 microns. The voltages required for the symmetrical four-pole groups used to produce the astigmatic beam would be about $\Phi_A/20$ (Archard 1955 a).

§ 9. EFFECT OF A CORRECTIVE LENS ON THE 'OTHER RAY'

The arrangement of the correction systems described above is such that the ray *not* under correction goes through the centre of the corrective lens, and is not affected thereby. In practice, however, this ray will not have negligible height throughout the region of influence of the corrective lens, and it is possible that an undesired aberration may be introduced into it. It is therefore necessary to evaluate approximately the aberration caused, for example, to the y_α ray by the x_α corrector. It will be sufficient to consider this for Scherzer's unperturbed corrective lens, as approximated by a combination of three two-pole groups.

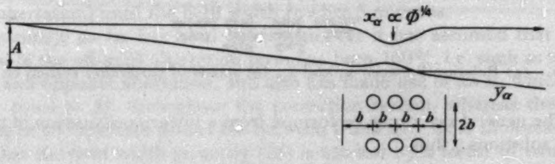


Figure 5. Effect of x_α corrector on y_α ray.

If the value of Φ_2 given by (17) be inserted in the ray equation for the yz plane (15) instead of in (4) (the equation for the xz plane) as before, it may be shown that an unfavourable image displacement arises corresponding to

$$C_c=\frac{29\pi}{32}\frac{\bar{A}^2v^2}{b} \qquad \text{......(23)}$$

where $\bar{A}$ is the height of the ray at a distance b from the corrective lens in question. It is desired that (23) shall not be appreciable in comparison with the v^2 part of the C_c given by (13) (for this is the operative part when two two-pole groups are put together). This requires that $\bar{A}/A$ be very much less than 2 which is certainly always the case as may be judged from figure 5 ($\bar{A}/A \sim \frac{1}{10}$ to $\frac{1}{5}$). Thus, the unfavourable C_c given to the y_α by the x_α corrector is negligible compared with the favourable C_c which will subsequently be imparted by the y_α corrector.

§ 10. OFF-AXIAL CHROMATIC CORRECTION

The image displacements which were given in (12) and (16) related to points on the axis of the beam. For extra-axial points, the chromatic conditions are more rigorous. According to Scherzer (1947), the image displacement corresponding to an object point $x=x_0$, $y=0$ is, for a unipotential lens,

$$[x_1-Mx_0]=\frac{Mx_0\epsilon}{\sqrt{\Phi_A}}\int_{z_0}^{z_1}\left(\frac{3}{8}\frac{\Phi'^2}{\Phi^2}-\frac{\Phi_2}{\Phi}\right)\frac{(\alpha x_\alpha{}^2+x_0x_\gamma x_\alpha)}{\sqrt{\Phi}}\,dz \qquad \ldots\ldots(24)$$

which consists of two parts: one, containing $x_\alpha{}^2$, is the same as the axial aberration, and is made zero by the correction system; the other, containing $x_\gamma x_\alpha$, is in general not made zero.

It is convenient to recall the significance of x_γ. Whereas x_α is the particular solution of the paraxial ray equation with object plane conditions $x_{\alpha 0}=0$, $x_{\alpha 0}'=1$, the independent solution x_γ has $x_{\gamma 0}=1$ but, as yet, has no restriction on its initial slope. The general ray x may be formed from a linear combination of these two particular solutions; thus

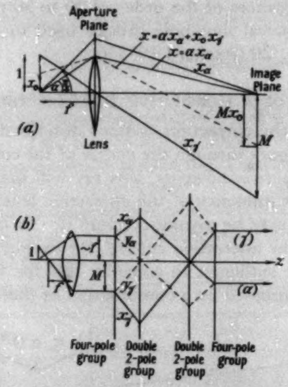


Figure 6. (*a*) Relation between x_α and x_γ, (*b*) action of correction system on x_γ, x_γ.

$$x=\alpha x_\alpha+x_0x_\gamma \qquad \ldots\ldots(25)$$

represents a ray starting from the object point $x=x_0$, $y=0$, with a slope depending on the initial slope assigned to x_γ. If (as is often chosen) x_γ is made to vanish in the aperture plane, then the x of (25) characterizes the ray which will meet the aperture plane at the same point as the ray $x=\alpha x_\alpha$ which starts from the axial point of the axis. It thus characterizes the most widely diverging ray which starts from the point $x=x_0$, $y=0$, and yet passes through the aperture. Further, in the gaussian image plane, which lies at the second zero of x_α, the value of x_γ is M, whatever its initial slope may be. A simplified geometrical picture appears in figure 6 (*a*). Here, x_γ passes cleanly through the lens centre (which is the same as saying that it becomes zero in the aperture plane) travelling right to the image plane in a straight line.

In order to simplify the estimation of off-axial chromatic aberration when the axial correction conditions have been fulfilled, consider another x_γ, this time passing through the back focus of the lens under correction and hence running from the lens to the image plane at height M (figure 6 (*b*)). Then, if this is regarded as a separate ray, the correction system will act on it in very nearly the same way as on x_α, with the scaling factor M/f. This is indicated in figure 6 (*b*), in which x_α and x_γ have been scaled differently because strictly, if x_α were to emerge from the lens exactly parallel to the axis, M would be infinite.

If now the essentially non-axial part of the chromatic aberration in the uncorrected lens were equal to M/f times the axial part, thus

$$\int \frac{3}{8}\frac{\Phi'^2}{\Phi^2}\frac{x_\alpha x_\gamma}{\sqrt{\Phi}}\,dz = \frac{M}{f}\int \frac{3}{8}\frac{\Phi'^2}{\Phi^2}\frac{x_\alpha{}^2}{\sqrt{\Phi}}\,dz \qquad \ldots\ldots(26)$$

(integrals taken over the lens under correction), the off-axial aberration would be corrected as well as the axial aberration. But this condition clearly does not hold, for x_γ crosses the axis in the neighbourhood of the lens, while x_α does not. In fact, the off-axial aberration term (left-hand side of (26)) will be *over*-corrected† by the correction system, so that, in the limit where the over-correction is large compared with (26), the following relation will hold (using (24))

$$\frac{\text{residual off-axial aberration}}{\text{original axial aberration}} = \frac{M}{f}\frac{x_0}{\alpha}. \qquad \ldots\ldots(27)$$

Suppose it is desired that this should not exceed $\frac{1}{10}$ (i.e. that there should be everywhere at least a ten-fold fall in chromatic aberration). Then this will only hold for object points closer to the axis than x_0, where

$$x_0 = \alpha f/10M. \qquad \ldots\ldots(28)$$

Taking the reasonable values $\alpha = 0{\cdot}01$, $f = 0{\cdot}5$ cm, $M = 20$, this yields $x_0 = 2500$ Å which means that the residual chromatic aberration will be less than one-tenth of the original axial aberration over a field of $\frac{1}{2}$ micron, and that it will not become as great as the original axial aberration (still considerably less than the original off-axial aberration) until the field width reaches 5 microns.

The estimate so far has been pessimistic, for it has assumed that the over-correction of the off-axial aberration term has been 100%, i.e. such as to produce an equal and opposite aberration, and also has made use of an x_γ which remains constant, equal to M, throughout the correction system, whereas the actual x_γ (vanishing in the aperture plane) lies between 0 and M. It is thus reasonable to suggest that the field width given by (28) is too low by a factor of two or three, so that in fact the field over which the aberration is less than one tenth of the original axial aberration is rather more than 1 micron wide.

---

† The statement that the off-axis aberration component is over-corrected may be checked by reference to work of Glaser and Schiske (1954). The ratio of the off-axis aberration δl for a lens with a bell-shaped electrostatic potential distribution to the axial aberration δr is, in their notation,

$$\frac{\delta l}{\delta r} = \frac{x_0}{d\alpha}\frac{2z_0/d}{1+(z_0/d)^2}$$

where z_0 is the z of the object point, d is the parameter of the bell-shaped field. Typical values would be $d = z_0 = f$ so that $\delta l/\delta r = x_0/\alpha f$ compared with the ratio $Mx_0/\alpha f$ for the correction system. This may be seen in a general way from the figure 6 (*a*), for x_γ keeps in the neighbourhood of unity inside the lens but rises from 1 to M as it passes through the correction system to the right of the lens. For other values of z_0 and d, the ratio $\delta l/\delta r$ would alter slightly but, in general, would remain a factor of the order of M below the ratio for the correction system.

§ 11. CONCLUSIONS

It has been shown that a suitably arranged group of four electrodes (a combination of two two-pole groups) is capable of providing axial chromatic correction in one plane with or without deflection of the electron beam. This makes it available for use either in Scherzer's chromatic correction system (which requires no deflective action of the corrective group), or in a proposed new simplified chromatic correction system which has certain advantages over that of Scherzer, but which requires deflective as well as corrective action on the part of the pair of two-pole groups; it is shown that this requirement can be met, and curves are given (figure 4) to relate deflection and correction to the voltages on the electrodes of the group.

In order to correct a lens with chromatic aberration coefficient $C_c = 1$ cm, the voltages on the corrective groups would have to be of the order of one tenth of the accelerating voltage, the electrodes being laminae with opposite edges 0·5 cm apart, and extending parallel to the axis for 0·8 cm. The length of the correcting system would then be about 10 cm. In order to avoid aberrations (due to constructional inaccuracy) of more than 2 Å, the positioning tolerances of the electrodes would be about 20 microns.

Although this system would not correct chromatic aberrations of off-axial points, the residual aberration pertaining to these would be less than one tenth of the original axial aberration for a field width of rather more than one micron, the relationship between field and residual aberration being linear.

ACKNOWLEDGMENTS

The author wishes to thank Mr. M. E. Haine, Dr. G. Liebmann, Mr. T. Mulvey, Mr. A. W. Agar, and Mr. A. E. Ennos for helpful discussions and comment, and Dr. T. E. Allibone, Director of the Laboratory, for permission to publish this paper.

APPENDIX

One further point may eventually be of importance. As was noted in the Introduction, a great deal of attention has been paid to correction of spherical aberration and very little to chromatic correction. Until comparatively recently, both tended to be submerged by the more manifest image defect of first-order astigmatism. Means of removing the astigmatism have been developed, and it is found in practice that, of the errors remaining, spherical aberration is at present considerably less important than chromatic aberration. Consequently, the systems which have been developed for the correction of spherical aberration may not come into useful service until chromatic aberration has been fully overcome. It is well, however, to be prepared, and it would be particularly useful if spherical aberration could be corrected as well by the addition of some simple device to the chromatic correction system outlined above. That this could be done is seen on comparison with earlier work (Archard 1955 a), which describes a system for the correction of spherical aberration differing from the chromatic system above only in the replacement of the modified Scherzer correctors (combinations of two two-pole groups) by eight-pole groups. As these do not affect the chromatic integrals, and as the chromatic correctors do not adversely affect the spherical aberration integrals (for four-pole groups only enter to a secondary extent into spherical correction, which is mainly effected by eight-pole groups, as shown in the work quoted), a combined corrective system could be formed by simple superposition.

REFERENCES

ARCHARD, G. D., 1954, *Brit. J. Appl. Phys.*, **5**, 294; 1955 a, *Proc. Phys. Soc.* B, **68**, 156; 1955 b, *Proceedings of the International Conference on Electron Microscopy, London,* 1954 (London : Royal Microscopical Society), in the press.

BOERSCH, H., 1953, *Z. Phys.*, **134**, 156.

GABOR, D., 1944, *The Electron Microscope* (London : Hulton Press), Chapter 6; 1951, *Proc. Phys. Soc.* B, **64**, 244.

GLASER, W. and SCHISKE, P., 1954, *Optik*, **11**, 422.

HAINE, M. E., and HIRST, W., 1953, *Brit. J. Appl., Phys.*, **4**, 239.

KUSHNIR, YU. M., BIBERMANN, L. M., and LEVKIN, N. P., 1951, *Bull. Acad. Sci. U.S.S.R.* (Phys. Series), **15**, 303.

LEONHARD, F., 1954, *Z. Naturf.*, **9a**, 727, 1019.

RANG, O., 1949, *Optik*, **5**, 518.

SCHERZER, O., 1947, *Optik*, **2**, 114.

SEELIGER, R., 1951, *Optik*, **8**, 311.

The Low Temperature (11°-20°K) Specific Heat of α and β Manganese

By G. L. BOOTH, F. E. HOARE and B. T. MURPHY
The Physics Laboratories, The University, Leeds

MS. received 31*st May* 1955

Abstract. The specific heat of α Mn varies from 0·19 joule mole^{-1} deg^{-1} at 12°K to 0·49 joule mole^{-1} deg^{-1} at 20°K; the corresponding values for β Mn are 0·62 joule mole^{-1} deg^{-1} and 1·18 joule mole^{-1} deg^{-1}. A plot of C_v/T against T^2 for α Mn yields a straight line and a γ value, the electronic heat coefficient, of 28×10^{-4} cal mole^{-1} deg^{-2} (118×10^{-4} joule mole^{-1} deg^{-2}). In the case of β Mn a similar plot shows a marked curvature below about 15°K and hence does not allow of the derivation of a value for γ.

Measurements of the specific heat of manganese over the temperature range 14°K to 22°K have been reported by Elson, Grayson-Smith and Wilhelm (1940), and Armstrong and Grayson-Smith (1949). Kelley (1939) has measured the specific heat of α Mn from 53 to 290°K and Shomate (1945) has investigated the specific heats of α Mn and γ Mn over the same temperature range. The interest in low-temperature determinations arises from the possibility of separating the electronic and lattice contributions to the specific heat. This is given added interest as the experiments so far reported give a value between 35×10^{-4} and 44×10^{-4} cal mole^{-1} deg^{-1} for the coefficient γ in the term γT representing the electronic heat; this value is amongst the highest reported for metals.

For the present investigation spectrographically standardized manganese supplied by Johnson Matthey and Co. Ltd., Laboratory No. 4135, was used. The material contained traces only of magnesium and calcium. A specimen was heat treated by annealing at 1120°C, measured with a disappearing filament pyrometer, for 16 hours in an argon atmosphere followed by rapid water quenching. It had been hoped that this procedure would result in γ Mn (Potter and Lukens 1946) but subsequent X-ray examination by the powder method, using molybdenum radiation, showed the specimen to be in the β phase with a lattice constant of 6·312 Å. This value is to be compared with that of 6·302 Å (corrected) obtained by Preston (1928). The identification of the phase is quite unambiguous and the difference of lattice spacings may be due in part to the use of radiation giving somewhat lower angle reflections. Only one weak foreign line was observed on the photograph; this is probably to be attributed to MnO but the identification is uncertain.

The specific heat of this specimen, in the form of small irregular plates, was determined over the range 11°K to 20°K by sealing it in a copper calorimeter with helium at a pressure of 5 cm as an exchange gas. The mass employed was approximately 60 g.

The results obtained are shown in figure 1 as curve I. It will be seen that the agreement of the results obtained on 19th August and 20th August and 14th September, all on the same specimen without removal from the calorimeter, is very satisfactory. This indicates that the rate of transformation from β Mn to α Mn, the room temperature stable modification, is quite slow at ordinary temperatures.

The same figure shows the mean curve, II, obtained by Armstrong and Grayson-Smith (1949). Nothing is stated by these workers regarding the phase of the material used and our supposition was that the normal room temperature stable form, α Mn, had been employed. To test this supposition our specimen of β Mn was now heated to 1100°c followed by a slow cooling to room temperature, afterwards being held at 600°c for three hours followed by a slow cooling to room temperature taking a further four hours. The specific heat was again determined, giving the curve III of figure 1. X-ray analysis of this specimen identified it as α Mn with a lattice parameter of 8·914 Å from the higher order reflections in agreement with the (corrected) value of 8·912 Å obtained by Preston.

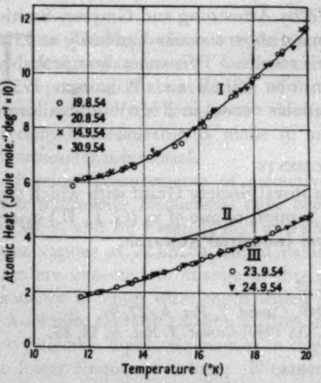


Figure 1. The atomic heats of α and β manganese; I, β Mn; II, mean values Armstrong and Grayson-Smith (1949); III, α Mn.

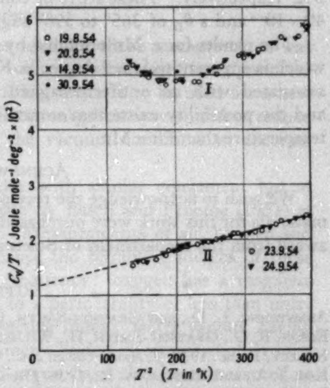


Figure 2. C_V/T against T^2 for α and β manganese. I, β Mn ; II, α Mn.

The large difference between curves I and III made it necessary to confirm the effectiveness of the heat treatment and quenching in reproducing the β phase as the x-ray work had not then been completed. The specimen was again transformed to the β form, by a similar treatment to that previously employed, and a few check specific heat measurements made. These confirmatory results are those for 30th September shown in figure 1.

As a check on the accuracy of the calorimetry the specific heat of copper was determined (20th September) over the same temperature range. The mean curve of Kok and Keesom (1936) gives an excellent representation of our results.

Discussion of Results

The position of the curve due to Armstrong and Grayson-Smith relative to our two curves suggests that the material used by them might not have been in a single phase. It should be remarked that it is not possible to obtain a good representation of their results by assuming a mixture of α Mn and β Mn and using our results for these. This is perhaps not surprising as manganese can exist in two further phases, γ and δ, beside α and β. The difference of approximately 2% between the results obtained by Elson *et al.* and Armstrong and Grayson-Smith may possibly be attributable to a similar cause.

The results used in plotting curve I, figure 1, are shown plotted as C_v/T against T^2 in curve I, figure 2. The upward movement of the values below about 15°K indicates that little reliance can be placed on the values of γ and θ_D, the Debye temperature, obtained from the linear portion of the curve above this temperature. The change of slope may be due to some physical transformation occurring in the neighbourhood of 15°K. No such behaviour is seen in the case of α Mn as shown by curve II, figure 2. Extrapolation and measurement of the slope of this curve yield values for γ and θ_D of 28×10^{-4} cal mole^{-1} deg^{-1} and 392° respectively. These are to be compared with a γ value between 35×10^{-4} and 40×10^{-4} and a θ_D of 365° to 390° suggested by Armstrong and Grayson-Smith.

The results for α Mn obtained by Shomate show a marked anomaly at 95°K which is apparent to lesser degree in Kelley's results. This anomaly is probably associated with an antiferromagnetic transition (Shull and Wilkinson 1953) and the possibility exists that something similar occurs in β Mn but at a lower temperature than for α Mn.

ACKNOWLEDGMENTS

We wish to acknowledge the receipt of a Royal Society Grant with which the materials for this work were purchased and support of one of us (G. L. B.) by an award from the Department of Scientific and Industrial Research.

REFERENCES

ARMSTRONG, L. D., and GRAYSON-SMITH, H., 1949, *Canad. J. Res.* A, **27**, 9.
ELSON, R. G., GRAYSON-SMITH, H., WILHELM, J. O., 1940, *Canad. J. Res.* A, **18**, 83.
KELLEY, K. K., 1939, *J. Amer. Chem. Soc.*, **61**, 203.
KOK, J. A., and KEESOM, W. H., 1936, *Physica*, **3**, 1035.
POTTER, E. V., and LUKENS, H. C., 1946. *Metals Technology, Technical Publication* No. 2032.
PRESTON, G. D., 1928, *Phil. Mag.*, **5**, 1198, 1207.
SHOMATE, C. H., 1945, *J. Chem. Phys.*, **13** 326.
SHULL, C. G., and WILKINSON, M. K., 1953, *Rev. Mod. Phys.*, **25**, 100.

The Refractive Indices of Water Vapour, Air, Oxygen, Nitrogen and Argon at 72 kMc/s

BY K. D. FROOME
Metrology Division, National Physical Laboratory, Teddington, Middx.

MS. received 25th May 1955

Abstract. The refractive indices of water vapour, air, oxygen, nitrogen and argon have been measured in a cavity resonator at a frequency of 72 kMc/s (corresponding to a wavelength of 4·16 mm). By comparing the values obtained at this frequency with those previously established at lower microwave frequencies it is found that the refractive indices for air, oxygen and water vapour are modified in much the manner to be expected from dipole theory, whereas those for nitrogen and argon remain unchanged.

---

§ 1. INTRODUCTION

THE 72 kMc/s microwave refractometer used for these measurements was designed primarily for association with a microwave interferometer operating at a wavelength of 4 mm. This device is to be used for a new determination of the free-space velocity of electromagnetic waves, and to reduce wavelength measurements made in air to the vacuum condition an accurate refractometer is essential.

Precise measurements of refractive index with similar equipment but at lower microwave frequencies have been made by Essen and Froome (1951) at 24 kMc/s, and by Essen (1953) at 9 kMc/s. Measurements of air, oxygen and water vapour at 72 kMc/s are of value because the refractive indices of these gases are expected to change at this high frequency: oxygen has a magnetic resonance at 5 mm wavelength which leads to a permeability of less than unity at $\lambda = 4$ mm, thereby causing a drop in refractive index; water vapour, having heavy electric dipole resonances in the millimetre region, increases slightly over the lower frequency values. Nitrogen and argon have also been measured at $\lambda = 4$ mm; as these gases are not expected to have a refractive index which changes with frequency they provide a means of confirming the satisfactory operation of the equipment.

§ 2. METHOD AND APPARATUS

The method employed was substantially the same as that for the measurements at lower frequencies, the resonant frequency of a cavity resonator being determined when evacuated and when filled with the gas under test. The refractive index is the ratio of these frequencies.

The cavity operated in the H_{017} mode, its internal dimensions being: length 1·497 cm, diameter 2·220 cm. A tuning plunger (diameter 0·122 cm) of non-rotating design was again incorporated for the purpose of making small changes in resonant frequency.

The resonator was constructed from invar and internally copper-plated. As before it was used in a waveguide bridge circuit in order to obtain the maximum possible sensitivity. The coupling hole to the cavity was carefully adjusted until it was a nearly perfect 'match' to the waveguide feed. In these conditions it had a Q-factor of about 12 000 and with the sensitive detector described below a plunger setting on resonance could, with steady conditions, be made with a reproducibility to ± 4 parts in 10^8.

The source of millimetre waves was a 36 kMc/s stabilized klystron oscillator coupled to a silicon crystal harmonic generator tuned to maximum efficiency at 72 kMc/s. The very low output from this device necessitated a superheterodyne detector involving (for local oscillator) a second stabilized 36 kMc/s oscillator and harmonic generator, giving a 60 Mc/s intermediate frequency when 'mixed' with the primary signal.

Frequency measurements were made at 36 kMc/s in much the same manner as in the earlier work at 24 kMc/s, the principal difference being the addition of a 600 Mc/s stage to the frequency multiplier. The precision of frequency measurement was again at least as close as ± 1 part in 10^8.

§ 3. RESULTS

The results, expressed in the form $(n-1)10^6$, where n is refractive index, are given in the table together with those previously obtained at 24 kMc/s and 9 kMc/s.

Refractivities of Water Vapour and Dry Gases
(Dry gases at 0°c, 760 mm Hg, water vapour at 20°c, 10 mm Hg)

Gas	9 kMc/s	24 kMc/s	72 kMc/s	f_{av} (kMc/s)†
Water vapour	60·7 ± 0·2	60·7 ± 0·2	61·0 ± 0·2	72·021 86
Dry air (CO_2-free)	288·10 ± 0·10	288·15 ± 0·10	287·66 ± 0·11	72·008 76
Oxygen	266·2 ± 0·2	266·4 ± 0·2	263·9 ± 0·2	72·010 08
Nitrogen	294·1 ± 0·1	294·1 ± 0·1	294·05 ± 0·11	72·008 06
Argon	—	277·8 ± 0·2	277·7 ± 0·2	72·009 34

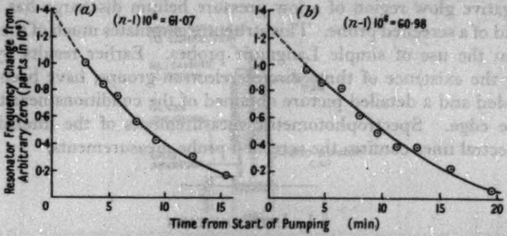
† f_{av} = average frequency of measurement for present results.

In all cases the error limits quoted represent the standard deviation of a single observation in statistical combination with estimated systematic errors.

The measurements on water vapour and the dry gases were made in exactly the same manner as those already described for the lower microwave frequencies, but for two reasons water vapour proved more troublesome than in the earlier work. Firstly, the presence of adsorbed molecules altered the electrical conditions on the interior surfaces of the cavity, producing a lowering of resonant frequency. Secondly, the heat of 'condensation' of these molecules caused an abrupt rise in temperature of between 0·1 and 0·2°c immediately upon admitting vapour, again producing a spurious lowering of resonant frequency.

A knowledge of the temperature coefficient of frequency of the cavity is all that is required to eliminate the second of these effects. The first can be eliminated by recording the resonant frequency as soon as possible after pumping out the vapour (i.e. as soon as a hard vacuum is obtained). Then further measurements are made at intervals of a few minutes until sufficient data are obtained to extrapolate to the instant that pumping started.

The figure shows two such curves made with the resonant frequency corrected to a constant cavity temperature. The dotted portions are the extrapolations. Referring to the time axis: the origin represents the start of pumping and it is seen that the first vacuum resonant frequency observation cannot be made until about three minutes have elapsed, this being the time required to obtain a good vacuum. For curve (*a*) the cavity was filled with about 12 mm Hg water vapour and left standing with this present for about 20 minutes before pumping out. For curve (*b*) the initial pressure was about 15 mm Hg which had been maintained in the cavity for two hours. There is thus some evidence that the longer the cavity is allowed to 'soak', the longer the adsorbed molecules take to be released. However, this conclusion is by no means established, for very considerable variation of curve shape (with apparently identical conditions) was obtained in the course of the experiments. On the graphs are shown the values of refractive index, corrected to the standard conditions, for the particular experiments represented.



Change of cavity resonant frequency with removal of occluded (or adsorbed) water molecules.

It is interesting to note that failure to correct for either of these disturbing effects gave rise to values of $n-1$, about 1% higher than the true value.

For a number of reasons neither of these disturbances was noticed at 9 kMc/s or 24 kMc/s. Firstly, the physical dimensions of these lower frequency cavities were much greater than those of the cavity operating at 72 kMc/s, so that the fractional dimension change produced by adsorbed or occluded molecules would have been correspondingly less. Secondly, the greater ratio of thermal capacity to cavity surface area would render the heat of 'condensation' effect smaller for the bigger cavities. Also, the 9 kMc/s and 24 kMc/s cavities had frequency–temperature coefficients of about 1×10^{-6} deg^{-1}, whereas the 72 kMc/s cavity had a rather high coefficient, namely 3×10^{-6} deg^{-1}.

From the foregoing remarks it would appear that considerable care would be needed in measurements of water vapour at radio frequencies where the sensitive element consists of a condenser of small plate separation. This could afford a reason why such determinations generally seem to give higher values than those reported herein.

ACKNOWLEDGMENTS

The author desires to acknowledge the assistance of Mrs. P. Anderton, who made most of the experimental observations. The work described above has been carried out as part of the research programme of the National Physical Laboratory and this paper is published by permission of the Director of the Laboratory.

REFERENCES

Essen, L., 1953, *Proc. Phys. Soc.* B, **66**, 189.
Essen, L., and Froome, K. D., 1951, *Proc. Phys. Soc.* B, **64**, 862.

Screened Probe Measurements in the Helium Negative Glow

By D. H. Pringle† and W. E. J. Farvis‡
† Ferranti Ltd., Edinburgh
‡ University of Edinburgh

MS. received 20th January 1955 *and in revised form* 15*th May* 1955

Abstract. Detailed examination of the electron energy and density distribution in the negative glow region of a low pressure helium discharge has been made with the aid of a screened probe. This structure eliminates much of the inaccuracy inherent in the use of simple Langmuir probes. Earlier results by Emeléus, indicating the existence of three discrete electron groups, have been confirmed and extended and a detailed picture obtained of the conditions near the cathode dark space edge. Spectrophotometric measurements of the intensities of the helium spectral lines confirm the screened probe measurements.

§ 1. Introduction

The conditions in the negative glow region of a helium discharge at pressures of the order of 1 mm of mercury were investigated by Emeléus and his group some years ago and were the subject of several papers (Emeléus 1927, Emeléus and Harris 1927, Emeléus and Brown 1929, Emeléus, Brown and Cowan 1934, Emeléus, Greeves and Montgomery 1937). In their experiments simple Langmuir probes were employed and information was obtained from the resulting semi-logarithmic plots of probe current against voltage about the electron energy and density distributions in the glow. They interpreted the results as being due to the existence of three Maxwellian groups of electrons with widely differing values of electron temperature and density.

An undesirable feature of the analysis of the results obtained with a Langmuir probe is that it is necessary to extrapolate the higher energy part of the semi-logarithmic plot to lower energies in order to separate the true component from the positive ion component. This leads to considerable inaccuracy in the detection of the higher energy electron component. In order to obtain more reliable information about the conditions in the negative glow the authors have employed screened probes, basically similar to that developed by Boyd (1950 a) but differing in constructional details. With this type of probe it is possible to measure the electron, positive ion and secondary emission components of the probe current directly.

As an important means of confirming the validity of the information obtained from the probe analysis, comparison has been made between spectrophotometric measurements of the relative intensities of the helium spectral lines and values calculated from the screened probe data. The latter calculations make use of excitation probability values of the helium lines as functions of electron energy as deduced by Lees (1932). The results are discussed in §3 of this paper.

A brief account of this work has been given by Pringle and Farvis (1954).

§ 2. Description of the Apparatus

Two discharge tubes, designated A and B, were employed. In both cases the electrodes and probes were capable of being moved magnetically, rotational movement as well as radial movement of the probes being possible. The electrodes in tube A were 2·7 cm in diameter and those in tube B are 6·4 cm in diameter, the maximum electrode spacing in each case being about 18 cm. In both tubes a Langmuir and a screened probe were mounted diametrically opposite each other at the centre of the tube.

A sketch of the screened probe is given in figure 1. The grid consisted of 0·0008 in. diameter tungsten wires spaced at 250 per inch and stretched across a 0·046 in. diameter hole in a Kovar tube, the tungsten wires being copper brazed to the tube. A thin nickel disc behind this acted as a collecting electrode, and in the experiments reported below this electrode was maintained at a potential of +80 volts with respect to the anode of the discharge tube. The leads to the grid and the collector consisted of the Kovar tube itself and a fine wire passing up the centre of an insulating sleeve. The Langmuir probe was of conventional design with the exposed tungsten collecting wire 0·64 cm long and 0·013 cm in diameter.

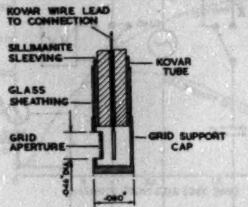


Figure 1. Screened probe assembly.

The discharge tubes were baked under vacuum to a temperature of 450°c before gas filling. Considerable care was taken to achieve a high degree of gas purity, and to eliminate the problem of contamination by mercury from the calibrating McLeod gauge a diaphragm manometer similar to that described by Alpert, Matland and McCoubrey (1951) was incorporated, but with the improved method of construction described by Pringle and Kidd (1953).

A schematic diagram of the circuit used to measure the probe characteristics, adaptable to both the Langmuir and Druyvesteyn methods of analysis, is shown in figure 2. In the Druyvesteyn method a 1000 c/s a.c. signal derived from a beat frequency oscillator was injected between the anode of the discharge tube and the grid of the screened probe through a simple changeover switch K and a large blocking capacitor C. The total collector current was metered and, by means of a balancing circuit, a sensitive galvanometer with a full scale deflection of 19 cm for 10^{-7} A recorded the increase of current Δi on injecting the 1000 c/s signal.

§ 3. Experimental Results

3.1. *Langmuir Analysis of Screened Probe Data*

By plotting the semi-logarithmic curve of electron current against grid voltage at varying positions in the negative glow it was found that one could resolve it into three distinct regions which were approximately linear. The fastest or primary group was however only detectable with the probe grid facing the cathode, indicating the existence of a fast directed stream of electrons coming from the cathode. The analysis of Langmuir and Mott-Smith (1924) can be applied to the determination of the space potential, temperature and concentration of the two slower groups which have a small ratio of drift velocity to random velocity except very near the cathode dark space edge. The analysis is however not applicable to the primary group in which the random isotropic component of electron velocity is negligible (Heatley 1937). Some remarks will be made later concerning the interpretation of the linear semi-logarithmic plot for the fast primary electrons.

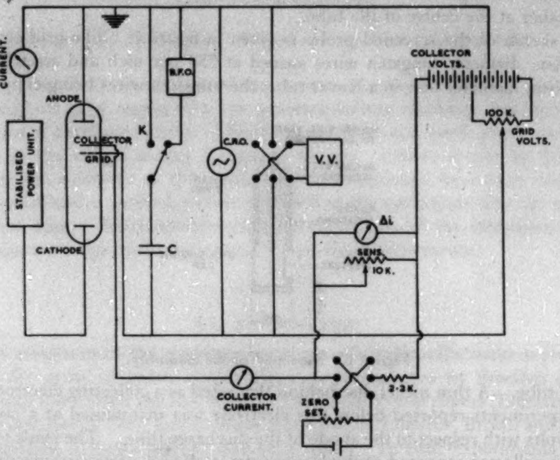


Figure 2. Probe circuit.

3.1.1. *Axial Distribution.*

The results presented in this section are for a 600 μA discharge in helium at a pressure of 0·58 mm of mercury, the anode–cathode distance being 6·4 cm. The negative glow was bluish-white in colour and extended from the cathode dark space edge to a region not far from the anode. The cathode dark space was dark green in colour, extending over about 13 mm and a faint pinkish cathode glow was present within 1 mm of the cathode.

The axial variation in the temperature and concentration of the two slower groups with the screened probe facing the cathode is given in figure 3. On moving into the negative glow from the edge of the cathode dark space (referred to below as C.D.S.) it will be noticed that the temperatures of both the slowest or ultimate group and that of the intermediate or secondary group fall sharply. There appears to be a minimum value of ultimate temperature near the brightest part of the glow. The variations of space potential and of secondary emission current with axial distance in the glow are reproduced in figure 4. Near the visual edge the exact height of the maximum in potential is indeterminate owing to the small change in curvature of the probe characteristic near space potential in that region, where the drift velocity is considerable (Heatley 1937).

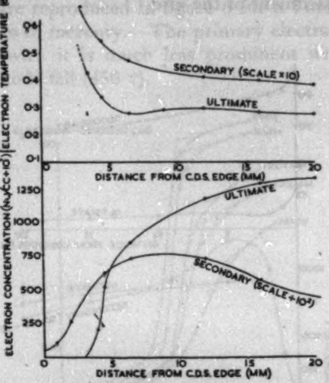


Figure 3. Variation of temperature and concentration of groups with axial distance (grid facing cathode).

Temperature and concentration variation of the ultimate and secondary groups throughout the negative glow with the screened probe grid facing the anode are given in figure 5. It will be seen at once that both the ultimate and secondary concentrations near the C.D.S. edge are larger than for the case of the probe facing the cathode, indicating the presence of a considerable electron drift in this region maintained by both the concentration gradient and the electric field. There would appear to be some electrons from these two slower groups present even on the C.D.S. side of the edge, although there is doubt about the validity of the measurements in this region owing to the large field intensities.

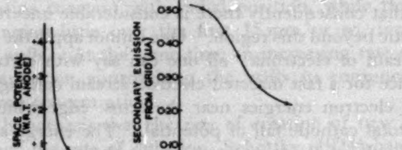


Figure 4. Variation of space potential and secondary grid emission with axial distance (grid facing anode).

No primary electrons were detectable anywhere in the negative glow with the probe facing the anode: the secondary group continued linearly in the semi-logarithmic plot down to currents of less than 0·01 μA. The directivity of the primary group is supported by another observation. This is that the secondary emission current from the probe was only about 0·07 μA, varying little with axial position, when the probe grid faced the anode, whereas it was much higher if the probe faced the cathode, decreasing with penetration into the glow (figure 4). The difference in secondary emission must be due to the shielding of the probe grid from the fast primary stream in the former case, with consequent reduction in positive ion bombardment of the grid.

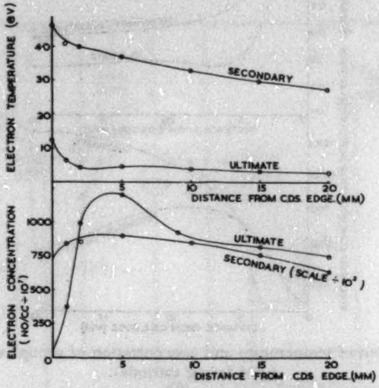


Figure 5. Variation of temperature and concentration of groups with axial distance (grid facing anode).

It must be realized that for retarding potentials greater than about 60 volts the secondary emission component is large compared with the primary electron component and that consequently there is considerable uncertainty in the shape of the characteristic beyond this region. One cannot apply the Langmuir analysis to a directed stream of electrons; all one can say with certainty is that there is definite evidence for a fast directed electron stream coming from the cathode dark space with electron energies near the C.D.S. edge which are appreciable fractions of the total cathode fall of potential. The energy and density of this primary beam fall off rapidly on penetrating into the negative glow.

Brief mention may be made of measurements with the unscreened Langmuir probe mounted diametrically opposite the screened probe in the discharge tube. These gave similar values of secondary and ultimate group temperatures and concentrations away from the C.D.S. edge to those obtained with the screened probe. The primary group could not be resolved with any certainty, however, even at comparatively low values of retarding potential, indicating the advantages of the screened probe structure. It may be noted that, as a test of the disturbance of the plasma by the screened probe, the potential of the grid was varied over a range from -100 to $+3$ volts with respect to the plasma potential and the effect on the Langmuir probe current observed, with the Langmuir probe situated about two millimetres distant. The maximum change in Langmuir probe current noted was not greater than 2%, implying that the screened probe potential does not disturb the plasma to any appreciable extent.

3.1.2. *Radial Distribution.*

In order to obtain a more accurate picture of the radial density and energy distribution of the electrons, measurements were made in the larger diameter tube B. Typical semi-logarithmic plots of electron current against grid voltage in the negative glow are reproduced in figure 6 for a 1 mA discharge in helium at a pressure of 0·7 mm of mercury. The primary electron contribution, while detectable, is not shown; it is much less prominent with the lower current density and lower cathode fall (450 v).

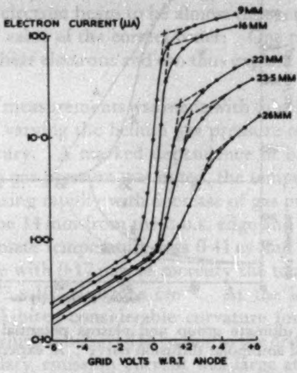


Figure 6. Semi-logarithmic plots for varying radial positions (distances marked are measured from centre of tube).

It is evident at once from inspection of the curves that the secondary group temperature remains constant with radial position, while that of the ultimate group, after changing little over the first 15 mm or so, increases sharply on approaching the walls. At the same time an increasing rate of change of space potential is apparent on approaching the walls in conjunction with a rapid decrease in electron concentration.

Now on the hypothesis that the rate of creation of new electrons per unit volume must equal the rate of diffusion, Schottky (1924) has shown that if the ionizing coefficient is assumed to be independent of spatial coordinates, implying no radial variation in energy distribution, the solution of the ambipolar diffusion equation may be written $n_r = n_0 J_0(2{\cdot}405r/R)$ where n_r is the electron density at radius r and n_0 is that at the tube axis. Taking $n_0 = 7 \times 10^9$ electrons cm^{-3} and $R = 3{\cdot}2$ cm we obtain the theoretical radial density distribution shown in figure 7 on this basis and this can be compared with the experimentally determined radial density distribution for the ultimate group using screened probe measurements. It will be noted that at the radius at which the ultimate group temperature begins to increase (about 15 mm) the density distribution begins to deviate markedly from the theoretical ambipolar diffusion curve. The plasma potential V_r at radius r from the axis can be written as

$$V_r - V_0 = \frac{kT}{e} \log \frac{n_r}{n_0}$$

if ambipolar diffusion theory applies. Taking $n_0 = 7 \times 10^9$ electrons cm^{-3}, $kT = 0{\cdot}27$ ev, the mean electron energy out to 15 mm radius, and $V_0 = +0{\cdot}6$ v with respect to anode, values of V_r for varying r have been calculated and are given together with experimental values in the lower half of figure 7. Again the curves deviate markedly beyond 15 mm radius.

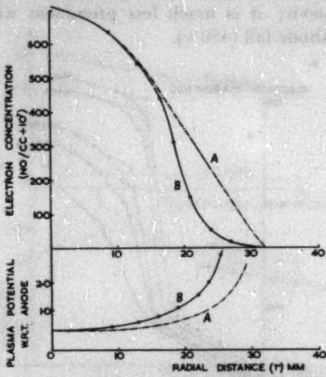


Figure 7. Density of ultimate group and plasma potential as function of radius. A, theoretical ambipolar diffusion curve ; B, experimental curve.

We can conclude from these results that the radial variation of electron concentration and potential are characteristic of ambipolar diffusion up to about half the tube radius from the axis but that beyond this region the increase in electron temperature causes considerable deviation. The radial distribution for cumulative ionization processes, discussed by Howe (1953), gives a curve not very different from the ambipolar diffusion curve. For radii less than half the tube radius, however, the ambipolar diffusion distribution gives a better fit with the experimental points.

3.1.3. *Variation with Discharge Current and Gas Pressure.*

Measurements of the effect on the electron group temperature and density of varying discharge current were made in tube A using helium gas at a pressure of 0·50 mm of mercury. The distance between the probe, which faced the anode, and the C.D.S. edge was maintained at 6·5 mm, the probe penetration being 10 mm.

(1)	(2)	(3*a*)	(3*b*)	(4*a*)	(4*b*)	(5)	(6)
600	1060	3·35	0·332	8·26	10·55	+1·05	0·06
500	980	3·35	0·304	7·25	9·40	+1·09	0·06
400	685	3·35	0·267	6·15	8·40	+1·06	0·05
300	530	3·35	0·232	4·87	5·88	+1·03	0·04
200	440	3·35	0·207	3·88	3·33	+1·07	0·03

(1) Discharge current (μA); (2) voltage drop; (3) electron temperature (ev) (*a*) secondary, (*b*) ultimate; (4) electron concentration (*a*) secondary ($\times 10\,\mathrm{cm}^{-3}$), (*b*) ultimate ($\times 10^9\,\mathrm{cm}^{-3}$); (5) plasma potential (v w.r.t. anode); (6) secondary emission current (μA).

The table indicates that the secondary group temperature is independent of discharge current while that of the ultimate group falls steadily with current. There is a rough proportionality between the density of the ultimate group and the tube current, as might be expected.

Emeléus, Greeves and Montgomery (1937) also observed this independence of secondary electron temperature on the discharge conditions in helium, although they estimated the electron temperature to be of the order of 5 ev. They explained the constancy of the secondary group temperature in terms of collision theory; from this one would expect the mean energy of electrons liberated in ionizing collisions with a fast electron beam to be almost independent of incident electron energy and to have a value of the correct order. One can identify the secondary electron group with these electrons and can thus explain the observed temperature constancy.

A second series of measurements was made with tube B, maintaining a discharge current of 2 mA and varying the helium gas pressure over a range between 0·17 and 1·5 mm of mercury. A marked dependence of both ultimate temperature and concentration on gas pressure was noted, the temperature increasing and the concentration decreasing rapidly with decrease of gas pressure. To take specific values, with the probe 14 mm from the C.D.S. edge and the gas pressure 0·83 mm of mercury, the ultimate temperature was 0·41 ev and the concentration 6×10^9 electrons cm^{-3}, while with 0·17 mm of mercury the temperature was 0·74 ev and the concentration $4{\cdot}3 \times 10^8$ electrons cm^{-3}. At the lower pressures the semi-logarithmic plots exhibited considerable curvature for retarding potentials less than about 10 volts, which is believed to be due to the drift velocity of the secondary group. The secondary emission current was large at the lower pressures and depended markedly on grid orientation, indicating that the fast primary group was still present although this was not obvious from the semi-logarithmic plots. It is also noteworthy that if the discharge current density was reduced at the lower pressures, so that the anode–cathode potential was nearer the normal cathode fall, the semi-logarithmic plot showed distinctly less curvature.

3.2. *Druyvesteyn Analysis of Screened Probe Data*

Instead of assuming a Maxwellian distribution of velocities, as in the Langmuir method, it is possible to measure the electron velocity or energy distribution directly using a method due to Sloane and Macgregor (1934) which employs the general formula of Druyvesteyn (1930) for the velocity distribution. In practical units this equation may be written in energy form as

$$N(E)dE = \frac{0{\cdot}52 \times 10^{11}}{A}\left(\frac{8mE}{e^3}\right)^{1/2}\frac{d^2i}{dE^2}\,dE$$

where $N(E)dE$ represents the number of electrons with energies between E and $E+dE$, A is the probe area in square centimetres and i the probe current in amperes.

In the method due to Sloane and Macgregor a small a.c. signal is injected into the probe circuit and the resultant increase in d.c. current Δi gives the second derivative of current with respect to voltage directly. This method was applied to the determination of the electron energy distribution in the negative glow of a 600 μA discharge in tube A with the helium gas pressure 0·50 mm of mercury.

Results are discussed below for two different amplitudes of a.c. signal, 0·1 and 0·3 volt, with the frequency 1000 c/s. The screened probe grid faced the cathode 5 mm from the c.d.s. edge.

The Langmuir analysis gave the space potential as +0·45 volt with respect to the anode and the temperature and concentration of the ultimate group as 0·33 ev and $5{\cdot}7 \times 10^9$ electrons cm^{-3} respectively. The space potential is given in the Druyvesteyn analysis as the value of retarding voltage at which Δi becomes zero; this is +0·17 volt for the 0·1 v a.c. signal and +0·18 volt for the 0·3 v signal.

Figure 8 gives the energy distribution for two values of a.c. injection signal using the Druyvesteyn analysis. It will be noted that only the ultimate group is detectable with this method, the values of Δi being too small ($<0{\cdot}001\,\mu\text{A}$) to be detectable at values of electron energy corresponding to the secondary and primary groups. For energies greater than about 0·75 ev the distributions given by the two input signals are similar but at lower energies a marked deviation is evident. That for the 0·1 volt signal appears to be roughly Maxwellian in form with a peak at 0·17 ev; on the same graph a theoretical Maxwellian distribution is drawn with $kT = 0{\cdot}34$ ev and $n_0 = 1{\cdot}1 \times 10^9$ electrons cm^{-3}, the value of density given by calculation from the peak level ($1{\cdot}49 \times 10^9$) on the basis of a Maxwellian distribution. It would appear that except at energies above 1 ev, where there are rather more electrons in the theoretical Maxwellian distribution, the curve for the 0·1v signal and the theoretical curve are closely similar. It is also important to observe that the Langmuir and Druyvesteyn methods of analysis give very nearly the same values of electron temperature, 0·33 ev and 0·34 ev respectively. The discrepancy in density is discussed below.

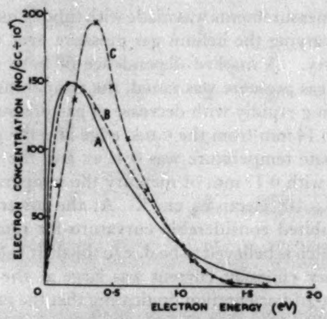


Figure 8. Druyvesteyn analysis.
A, theoretical Maxwellian distribution ; B, Druyvesteyn analysis with a.c. input=0·1 v; C, Druyvesteyn analysis with a.c. input=0·3 v.

The distortion introduced by the injection of an a.c. signal whose amplitude is appreciable in relation to the total width of the distribution is exemplified in the distribution curve based on the 0·3 v injection signal. This indicates the importance of keeping the a.c. injection signal small compared with the width of the distribution curve and of being able to detect the correspondingly small steady current increments.

It must be observed that the criterion for the determination of the space or plasma potential is quite different in the Druyvesteyn analysis from that in the Langmuir analysis, which is that the space potential is given by the intersection of the extrapolated saturation region of the semi-logarithmic plot and the extrapolated ultimate region. The Druyvesteyn criterion is equivalent to the determination of the value of retarding potential at which d^2i/dE^2 changes sign and becomes negative. On applying this criterion to the semi-logarithmic current–voltage plot it was found that the break-away point is indeed very near the $+0{\cdot}17$ v space potential level given by the Druyvesteyn analysis. For this value of space potential the Langmuir analysis gives an electron density of 2×10^9 electrons cm^{-3}, which is at least of the same order of magnitude as that given by the Druyvesteyn analysis ($1{\cdot}1\times10^9$ electrons cm^{-3}).

We can conclude then that the Druyvesteyn analysis of the ultimate group indicates an approximately Maxwellian distribution though showing rather fewer high energy electrons. This result is in accordance with earlier observations by Emeléus, Greeves and Montgomery (1937). If account is taken of the difference in space potential criteria the electron density agrees reasonably well with that given by the Langmuir analysis of the semi-logarithmic plot. Only the ultimate group is detectable by the Druyvesteyn analysis, the sensitivity limitations being inherent in the method.

3.3. *Spectrophotometric Measurements*

Detailed examination of the spectral lines in the cathode dark space–negative glow region of a helium discharge in tube A (0·58 mm Hg, 600 μA) was made with a Hilger Littrow Spectrometer Type E492, and the line intensities were measured with the aid of a microphotometer. The discharge tube was mounted so that the spectrometer slit was situated symmetrically with respect to the tube axis. A series of exposures, each of duration 30 minutes, was taken on the same photographic plate at various positions along the tube axis in the cathode dark space and negative glow. Some thirteen prominent HeI lines were registered while no HeII lines or impurity lines were detectable. The line intensities rise to a maximum a few millimetres into the glow.

Now important confirmation of the validity of the screened probe measurements would be obtained if there were agreement between the values of line intensity as a function of axial position in the discharge obtained by spectrophotometric measurement and those obtained from the screened probe results in conjunction with the excitation probability data for helium (see, for example, Lees 1932). The probe results give the variation of the number and energy of the electrons with axial position in the glow, the discharge conditions being identical to those employed in the spectrophotometric measurements. Thus, we can find for each line the product of the atomic cross section for electrons of a given energy times the number of electrons of that energy, as a function of electron energy in each group. By measuring the areas under the curves and adding contributions for each group we can obtain functions at each point in the discharge which should be proportional to the total intensity of emission of particular lines. Calculations of this type were made using the atomic cross section curves for line 3888·6 Å (3^3P–2^3S) and line 4471·5 Å (4^3D–2^3P), the two most intense spectral lines. Practically the entire contribution to excitation throughout the glow is attributable to the secondary electron group and the contributions from the other two groups is ignored in the following calculations. The two sets of values for the two lines, obtained by means of a planimeter, are plotted in figure 9 (*a*), (*b*). The units are arbitrary, the scale having been adjusted to give the best fit with the spectrophotometric values of relative intensity.

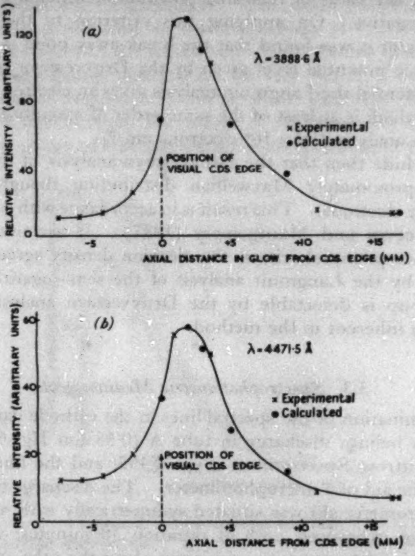


Figure 9. Variation of intensity with distance in glow, showing experimental spectroscopic values and values calculated from screened probe data and excitation probability values.

It will be observed that the points determined by the two different methods lie remarkably close to the mean curve, although there are not many spectrophotometric points in the negative glow. The peak of intensity in each case lies about 2 mm into the glow from the visual edge while the peak in total electron density lies about 6 mm into the glow, which is in qualitative agreement with the earlier observations of Emeléus and Harris (1927). The diffuse nature of the C.D.S. edge is evident in both curves, being more diffuse for the 4471·5 Å than for the 3888·6 Å curve, as one might expect owing to the less rapid change of excitation cross section with energy for the 4471·5 Å line.

Massey (1955) has reported recently that the theoretical atomic cross section curves for electrons in helium are much higher at low values of electron energy than Lees' experimental results would indicate. If this is true then it would be the slow ultimate group of electrons which would contribute chiefly to the spectral emission; it would then appear difficult to reconcile the probe measurements at the C.D.S. edge with the spectrophotometric measurements owing to the low ultimate density there. In addition it may be noted that Bates and Damgaard

(1949) have pointed out that Lees' assumption that only the 3^3P–2^3S transition for the 3^3P state is possible, is in error. In fact the 3^3P–3^3S transition is allowable and from the theoretical values of the transition probabilities Lees' figures can be corrected. The correction is, however, small and does not appreciably alter the figures employed here for atomic cross sections.

§ 4. Discussion of Results

Of the three groups of electrons detected by screened probe analysis in the negative glow, the interpretation of the existence of the fastest or primary group would appear to be the most straightforward. The directivity of this group has been clearly demonstrated and the electrons must have acquired their high velocities by acceleration in the very strong field in the cathode dark space region. Some of the energy gained by the field is, however, lost in collisions in the dark space, which emits a dark green radiation, presumably due to collisions involving fast electrons. The presence of a fast electron group was established in earlier work by Emeléus and his group using Langmuir probes but accurate measurements were not possible owing to the limitations of the method in the high energy region; Emeléus and Brown (1929) believed there was an isotropic fast electron group in the negative glow with an approximate Maxwellian distribution. Brewer and Westhaver (1937), however, employing quite a different experimental method, showed that primary electrons enter the negative glow from the cathode dark space with practically the full cathode fall of potential. In their experiments electrons coming from the dark space passed through a knife-edge slit in the discharge tube anode and were analysed magnetically with the aid of a fluorescent screen. As the negative glow was introduced across the slit some inhomogeneity of the beam was detected, but unfortunately no estimate was made of the energy distribution.

In addition to the fast primary group it has been confirmed, as earlier workers had observed, there would appear to be two slower groups in the negative glow, with nearly Maxwellian energy distributions. Emeléus and Brown (1929) tentatively advanced the suggestion that the existence of the two slow groups might be associated with the existence of a minimum in the Ramsauer electron free paths. While a sharp minimum does occur at about the mean energy of the secondary and ultimate groups for the gases investigated by the above workers, namely neon, argon, krypton and xenon at about 1 mm pressure, such a minimum in the case of helium, if present, is very small (Loeb 1939). Donahue and Diecke (1951) observed stationary and moving striations in the negative glow region of discharges with a positive column and put forward an explanation for the mechanism of their production. Such phenomena, which might have accounted for the presence of different energy groups of electrons, could not have been present in the negative glow experiments described here because of the absence of a positive column. The absence of such oscillations was verified experimentally.

A more likely explanation of the existence of two slow groups, as discussed by Emeléus (1951) for a low pressure mercury discharge, seems to be that the secondary group corresponds to those electrons which have been ejected after the collision of the fast primaries with gas atoms. These electrons have not had sufficient time to reach thermal equilibrium with the slowest or ultimate electron group, which is produced by the degeneration of the primary and secondary electrons in successive collisions to very low energy before removal by diffusion to the walls. In addition to the earlier remarks about the relative independence of the secondary group temperature on the discharge conditions, additional support for this picture is provided by the screened probe results for the axial electron distribution in the glow. The curves of figure 3 indicate that the secondary electrons appear first on entering the negative glow from the cathode dark space and the ultimate electrons are first detectable several millimetres further into the glow. This is presumably due to the fact that the primaries have to move some way into the glow in order to produce a copious supply of ultimate electrons by multiple collision processes.

It is more difficult to understand why the secondary group has an energy distribution which is approximately Maxwellian. It may be that some type of equilibrium is brought about between the secondary electrons and radiation quanta, and indeed it has been shown that the secondary electron group is almost entirely responsible for excitation in the negative glow. The ultimate electron group increases in temperature outside a central dense plasma-like region, but the temperature of the secondary group remains constant with radial position while the density changes slowly. This is reflected in the very slight decrease in light intensity in the glow on approaching the walls of the discharge tube.

ACKNOWLEDGMENTS

One of us (D. H. P.) wishes to express his indebtedness to Messrs. Ferranti Ltd., Edinburgh, for generously providing the facilities for carrying out these experiments, and to the Admiralty for sponsoring this research work.

REFERENCES

ALPERT, D., MATLAND, C. G., and MCCOUBREY, A. O., 1951, *Rev. Sci. Instrum.*, **22,** 370.

BATES, D. R., and DAMGAARD, A., 1949, *Phil. Trans. Roy. Soc.* A, **242,** 101.

BOYD, R. L. F., 1950 a, *Proc. Roy. Soc.* A, **201,** 329; 1950 b, *Nature, Lond.*, **165,** 228.

BREWER, A. K., and WESTHAVER, J. M., 1937, *J. Appl. Phys.*, **8,** 779.

DONAHUE, T., and DIECKE, G. H., 1951, *Phys. Rev.*, **81,** 248.

DRUYVESTEYN, M. J., 1930, *Z. Phys.*, **64,** 790.

EMELÉUS, K. G., 1927, *Proc. Camb. Phil. Soc.*, **23,** 531; 1951, *The Conduction of Electricity through Gases* (London: Methuen), p. 55.

EMELÉUS, K. G., and BROWN, W. L., 1929, *Phil. Mag.*, **7,** 17.

EMELÉUS, K. G., BROWN, W. L., and COWAN, H. M., 1934, *Phil. Mag.*, **17,** 146.

EMELÉUS, K. G., GREEVES, F. D., and MONTGOMERY, E., 1937, *Proc. Irish. Acad.*, **43,** 35.

EMELÉUS, K. G., and HARRIS, N. L., 1927, *Phil. Mag.*, **4,** 49.

HEATLEY, A. H., 1937, *Phys. Rev.*, **52,** 235.

HOWE, R. M., 1953, *J. Appl. Phys.*, **24,** 881.

LANGMUIR, I., and MOTT-SMITH, H., 1924, *Gen. Elect. Rev.*, **27,** 449.

LEES, J. H., 1932, *Proc. Roy. Soc.* A, **137,** 173.

LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharges in Gases* (New York: John Wiley), p. 649.

MASSEY, H. S. W., 1955, *International Symposium on Gas Discharges*, April 1955 (Delft).

PRINGLE, D. H., and FARVIS, W. E. J., 1954, *Phys. Rev.*, **96,** 536.

PRINGLE, D. H., and KIDD, R. M., 1953, *Rev. Sci. Instrum.*, **24,** 877.

SCHOTTKY, W., 1924, *Phys. Z.*, **25,** 635.

SLOANE, R. H., and MACGREGOR, E. J., 1934, *Phil. Mag.*, **18,** 193.

The Thermal Effects Associated with the Magnetization of High-Coercivity Materials

By L. F. BATES and A. W. SIMPSON†
University of Nottingham

MS. received 29*th March* 1955 *and in final form on* 31*st May* 1955

Abstract. The heat changes accompanying magnetization processes in materials of high coercivity, over a field range of 4000 oersteds, have been studied by a new method. The results are in good agreement with the Stoner and Rhodes theory of magnetization processes in such materials.

§ 1. Introduction

The thermal changes accompanying the magnetization of soft ferromagnetic materials have been extensively investigated in recent years, but, owing to the limitations of known methods, no measurements have been made on high coercivity permanent-magnet materials. Such measurements are of particular interest as they may shed light on the origin of the high coercivity in these materials. Fundamentally, measurements of the temperature changes, and hence of the quantities of heat generated, are made as materials are taken step by step through a hysteresis cycle.

In the method used by Bates and his co-workers (cf. Bates and Marshall 1953), the specimen in the form of a long rod is mounted in a water cooled solenoid, and about twenty differential copper–constantan thermocouples attached to it. Each thermocouple is connected to a separate primary of a carefully designed toroidal transformer. A sensitive galvanometer is connected to a secondary winding. The sensitivity of the galvanometer is amplified by a factor of about 10 by reading the deflection through a telescope or, more recently, by means of a simple galvanometer amplifier (Bates and Sherry 1955). When a field change is made, the change in the flux linked with the thermocouple leads, etc., induces an unwanted e.m.f. in the secondary of the transformer. In order to overcome the unwanted deflection, which can be of the same order as the deflection produced by the temperature change, a compensating coil of wire is connected to another primary winding and placed near the solenoid. By varying the position of the coil relative to the solenoid the inductive pick-up voltage can be compensated. A different setting of the compensating coil is usually required for each point in the hysteresis cycle.

§ 2. Method

Since high coercivity materials are normally available only as short specimens of the order 3 to 4 cm long, and require much greater fields to magnetize them to saturation than are easily obtained with a solenoid, measurements must be made on such specimens mounted between the pole-pieces of an electromagnet. Attempts were made to use the multi-thermocouple method described above with an electromagnet, but they proved unsuccessful, largely owing to the big inductive pick-up voltage induced by the stray field of the magnet. It was impossible to compensate for the induced e.m.f. because the latter was not only large, but appeared to originate from two sources with different time variations, presumably as they were severally generated by the coils and the yoke, the eddy currents in the yoke delaying the stray field of the latter.

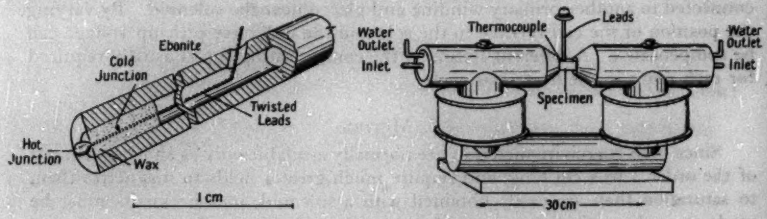
---

† Now with Plessey Co., Towcester, Northants.

A new method of measuring small temperature changes was first developed using the 'out of balance' voltage of an alternating current thermistor bridge. (A thermistor is a temperature-sensitive semiconducting resistance.) The change in out of balance voltage produced by the change in thermistor resistance as the temperature of one of the thermistors was changed, was amplified and displayed by a pen recorder. Unfortunately, the noise voltage generated by the bridge limited the sensitivity to a figure much lower than that required. The source of the noise voltage was traced to a random variation in the resistances of the thermistors, a result confirmed later by Doucet (1951).

The method eventually adopted consisted essentially of using a single differential thermocouple and a sensitive galvanometer with galvanometer amplifier. If the thermocouple leads are made from twisted wire, the finer the wire used the smaller will be the effective loop area to pick up the undesirable e.m.f. induced by a change in the field. On the other hand, the e.m.f. generated in the thermocouple by the temperature change is independent of the wire gauge, and therefore, by using fine, tightly twisted wire the ratio of signal to induced e.m.f. can be greatly increased. Unfortunately, by using fine wire the resistance of the thermocouple circuit is increased and the galvanometer deflection correspondingly decreased. Such reduction in sensitivity, and that involved by using only one thermocouple with no matching transformer, was overcome by using a very sensitive galvanometer amplifier.

The thermocouple was made from No. 36 s.w.g. copper and No. 37 s.w.g. constantan double silk covered wire. The cold junction was 2 or 3 mm from the hot junction, and the thermocouple and its leads were very tightly twisted together. The thermocouple was cemented along the axis of an ebonite tube, with the hot junction protruding from one end and bent over so as to rest on the flat surface of the end of the tube, as in figure 1 (*a*). The ebonite tube was mounted symmetrically with respect to the electromagnet, and the hot junction was pressed by a small spring on to the ground and polished surface of the specimen, the latter being trapped between the pole-pieces of the electromagnet,



(*a*) (*b*)

Figure 1. (*a*) Design of thermocouple. (*b*) Electromagnet and thermocouple (approximately to scale).

as shown in figure 1 (*b*). The ebonite tube was mounted so as to rotate about its axis, through a known angle, while the thermocouple made good thermal contact with the specimen. Any small induced e.m.f. that now remained could be completely eliminated by turning the ebonite tube.

For each material investigated, a region was found in the thermal cycle where the change in temperature produced by a change in applied field was small. Then, for a field change in this region, a graph of observed galvanometer deflection against the position of the thermocouple, measured by the angle through which the tube is turned from a datum position, was made for each direction of the applied field. Whereas the sign of the induced e.m.f. changed with change in field direction, that of the true thermoelectric e.m.f. did not; consequently two approximately linear curves were obtained, one for each direction of the field. The angle corresponding to the point of intersection of the two curves was the thermocouple setting for complete compensation, and the deflection at this point corresponded to the true thermal change. The setting for zero induced e.m.f. fortunately remained constant throughout the whole set of readings.

Although a differential thermocouple was used to reduce the effect of gradual changes in ambient temperature, it failed to overcome zero drift of the galvanometer. Apart from changes in room temperature, heat from the field coils of the electromagnet produced effects. Partly to reduce these effects, and partly to allow magnetostriction to take place freely during magnetization, thin sheets of soft rubber were placed between the pole-pieces and the specimen. The maximum thickness of rubber was limited by the permissible reduction in field introduced by the gap made in the magnetic circuit.

A special electromagnet, shown in figure 1 (*b*), was designed with its field coils as far away as practicable from the specimen and thermally insulated by a layer of poorly conducting material between the coil bobbins and the yoke. The ends of the pole-pieces were drilled out, and water from a constant temperature supply circulated through them. An effective 'thermal guard ring' was placed round the specimen by surrounding the gap by a metal container which was in good thermal contact with both pole-pieces. The region between the 'guard ring' and the specimen was packed with cotton wool. As a final precaution the whole electromagnet assembly was enclosed in a box, and the experiments were performed in a room with no windows or outside doors.

Special precautions were also taken with the galvanometer amplifier. A simple single stage galvanometer amplifier with no feedback was used with a Cambridge type L galvanometer for the primary galvanometer and a modified Tinsley type for the secondary galvanometer. A selenium barrier layer, split photoelectric cell was used. To improve the stability of the amplifier, it was mounted on a large slate slab resting on three pillars cemented to the foundations of the building. A considerable improvement in stability was obtained by placing a concave mirror behind the lamp so as to produce an image of the filament in the same plane and close to the actual filament (Jones 1951). This increased the illumination of the split cell and also reduced detrimental effects produced by slight movements of the filament. If the filament moved, its image moved an equal amount in the opposite direction with the result that the cell illumination remained symmetrical.

Although negative feedback was not used, the amplifier was linear to within 4% over the working range of the secondary galvanometer scale, but the sensitivity appeared to vary from day to day, even with stabilized lamp voltage etc. In order to allow for this variation, a calibrating circuit was made to produce three standard deflections, either positive or negative, and immediately before a thermal reading was taken, a calibration deflection was recorded of the same sign and approximately the same magnitude as that of the thermal deflection about to be measured. A further correction had to be applied to each reading, for despite such precautions as thermal lagging, there remained a slight drift of the galvanometer zero. The drift was therefore timed for a short period before a reading was taken and a small correction applied to the latter. The sensitivity of the whole system, though considerably less than that used by Bates and his co-workers, was adequate for the experiments, being at its highest 5400 erg cm^{-3} for 1 cm deflection. The stability was such that readings of the order of 0·5 cm could be made with reasonable accuracy. The mean of at least ten readings was taken for each measurement.

§ 3. RESULTS

Starting from a cyclic state, thermal readings were taken as the applied field was changed step by step round the hysteresis cycle. After the cycle had been traversed ten or more times, the algebraic sum of the mean of each reading for the whole cycle was found and equated to the area of the particular hysteresis cycle, calculated in ergs. From this was deduced the sensitivity of the system in ergs per cm deflection and hence the behaviour of the thermal changes with field. The more direct forms of calibration used by previous workers, including the method of electrical heating, were found unsatisfactory.

An eddy current correction was investigated by traversing a part of the hysteresis cycle, first in one complete step, then in two, three . . . up to, finally, the number of steps actually used in the thermal curve measurements. The thermal change observed when traversing this particular part of the cycle in one step is the sum of the real thermal change due to the change in magnetization plus a term due to eddy current heating. When the change is made in two steps the total genuine thermal change remains unaltered, but the eddy current heating is reduced. For a particular electromagnet it can easily be shown that the eddy current heating is approximately proportional to a parameter which is the product of the total electromagnet circuit resistance after the change is made and the square of the flux change involved. When the sum of the values of this parameter obtained when the cycle was traversed in one, two . . . steps, was plotted against the corresponding total thermal deflection, an approximately linear relation was obtained. The intercept for zero eddy current parameter gave the true total thermal deflection, while the slope of the line gave the constant of proportionality necessary for calculating the eddy current correction. For all the materials investigated the eddy current heating was in each case found to be negligible, and no correction was necessary. Presumably, the low eddy current heating was due to the large time constant of the electromagnet and associated circuit.

The (B, H) hysteresis curves were measured in the usual manner, a small magnetic potentiometer being used for the effective field measurements, and a coil of 10 to 20 turns of fine wire wound on the specimen for the induction measurements. Readings of both change in field and change in induction were obtained for the same steps in applied field as were made in the thermal measurements.

The thermal results obtained with all the materials were essentially of the same form. Figure 2 is typical; it shows the thermal curves for three different values of the maximum applied field (the largest being sufficient to saturate the material) for an industrially heat-treated specimen of Alnico. As the field was reduced from its maximum negative value every material investigated showed an initial cooling, as in figure 2, the variation being accurately linear with field over the first few hundred oersteds. This linear cooling continued on further reduction in the field in the case of some materials, but at a slightly lower rate for others, until in the neighbourhood of the positive coercive point there occurred a rapid increase in temperature, so that the bulk of the total heat generated throughout the course of a half cycle was released over a relatively narrow range of field. The rate of heating passed through a maximum and then decreased until, as saturation or maximum field was reached, its value was of the same order of magnitude as the initial cooling rate.

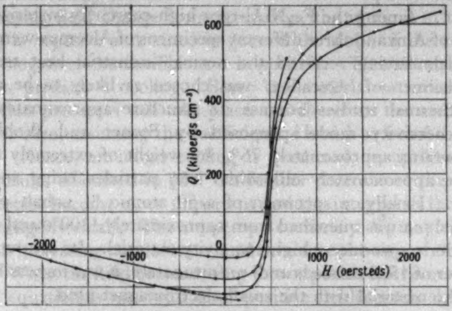


Figure 2. Thermal curves for Alnico.

Following Okamura (1936) attempts were made to separate the thermal readings into reversible and irreversible components for that part of the hysteresis cycle in which the field decreased from saturation to zero. This was achieved by traversing the cycle in the normal direction until a chosen applied field was reached and then making a field change in the opposite direction to that used previously. After each measurement the state of the material was again made cyclic.

If the magnetic process occurring in the region of the cycle over which the reverse field change is made is completely reversible, the thermal change obtained on making the reverse field change will be exactly equal but of opposite sign to that produced by a similar field change in the normal direction of description of the cycle. If, on the other hand, irreversible processes take place the forward and reverse thermal changes will differ in magnitude. The difference between the magnitude of the forward and the reverse thermal readings gives an approximate value of the irreversible component in the change of state under investigation. The condition for the difference between the forward and reverse thermal readings to be accurately equal to the irreversible component of the thermal change, is that the change shall be small. In the experiments described here this could not be so because of the comparatively low thermal sensitivity, and consequently the only real significance of the reversible readings is that they indicate over what region of this part of the cycle the magnetization processes are completely reversible. The reversible results for the Alnico specimen are indicated by the broken line in figure 2 and show that for initially saturated material the magnetization processes are completely reversible from saturation until the field is reduced below about 800 oersteds.

Experiments were performed on seven different materials of the compositions given in table 1. Among the Fe_2NiAl-type high-coercivity materials, a specimen of Alnico, one of Alni and three different specimens of Alcomax were investigated, all the materials having received the normal industrial heat treatments. In addition a specimen of 'Gecalloy' was chosen as likely to be an interesting material for thermal studies because its structure approximates very closely to the high-coercivity model proposed by Stoner and Wohlfarth (1948). 'Gecalloy' contains approximately 76% by weight of extremely finely divided pure iron, the approximately ellipsoidal iron particles being suspended in a plastic binder. Finally, a specimen of a 50 atomic % cobalt–platinum alloy was investigated; it was quenched from approximately 1300°c and then annealed at 750°c in order to produce a high-coercivity material. Its saturation coercivity was of the order of 1500 oersteds and, unfortunately, it was impossible to saturate (technically) the material with the small electromagnet used.

Table 1. Magnetic and Thermal Properties

Material	$H_{\max}$ (Oe)	I_{sat} (gauss)	$Q_{\max}$ (k-erg)	$H_{Q=0}$ (Oe)	$\left(\frac{dQ}{dH}\right)_{H_{\max}}$ erg Oe^{-1}	a
Alnico	2340	785	637	450	68·5	0·087
Alni	2370	620	594	570	48·0	0·077
Alcomax II	2710	930	915	470	108·5	0·117
Alcomax III	2540	815	1009	590	76·5	0·094
Alcomax IV	2480	1080	1401	635	96·0	0·089
'Gecalloy'	4200	245	267	545	21·8	0·089
Co–Pt alloy	4060	210	550	285	—	—

All the materials investigated except the cobalt–platinum alloy showed reversible behaviour similar to Alnico, the processes being completely reversible until decreasing fields between 800 oersteds and zero were reached. The cobalt–platinum alloy exhibited irreversible processes over the whole cycle, presumably because the material was not initially magnetized to saturation. The salient properties of the thermal curves are summarized in table 1. The field $H_{Q=0}$ is that at which the thermal curve cuts the field axis, and $(dQ/dH)_{H_{\max}}$ is the gradient of the curve obtained on reducing the field from its maximum value.

§ 4. Analysis and Interpretation of Results

In a complete thermodynamical treatment of the reversible processes accompanying magnetization, Stoner and Rhodes (1949) have shown that the reversible component of the thermal change ΔQ for a field change from H_1 to H is given by

$$\Big[\Delta Q\Big]_{H_1}^{H} = a\int_{H_1}^{H} d(HI) + b\int_{H_1}^{H} H\,dI \qquad \ldots\ldots(1)$$

where I is the magnetization and a and b are constants for the material under investigation. The first term in the expression represents the thermal change associated with the change in intrinsic magnetization, the constant a being equal to $-(T/I_0)(\partial I_0/\partial T)_H$, where T is the absolute temperature and I_0 the intrinsic magnetization. The partial differential coefficient is of course the slope of the magnetic corresponding states curve at the temperature of the experiment. The second term represents the thermal change associated with reversible rotational processes. The constant b is equal to $(T/K_1)(\partial K_1/\partial T)_H$, where K_1 is the appropriate anisotropy coefficient.

For a decreasing field change from saturation, the integral in the first term of equation (1) is negative and, since in general the slope of the corresponding states curve is negative, the coefficient a is positive. Therefore, the first term in the equation predicts a cooling. Since the material is initially saturated, negligible rotation processes will take place and therefore the second term will be very small or zero, assuming the coefficients a and b to be of the same order of magnitude. Hence, an initial cooling is predicted, in agreement with the experimental results discussed in § 3. On further decrease of the applied field, rotational processes become more important and the integral in the second term is no longer negligibly small. The anisotropy coefficient K_1 may either decrease or increase with increasing temperature, and therefore the sign of the coefficient b is not certain. When the field becomes very low or positive, irreversible processes occur and the equation is not valid until positive fields much above coercivity are reached.

Fe_2NiAl-*Type Alloys*

For the Alcomax II specimen the variations of Q, $\int d(HI)$ and $\int H\,dI$ with field, for field values where irreversible processes are small or zero, are shown in figure 3. The integrals were obtained by measuring the appropriate areas within the hysteresis loop. (For reasons given below, the axes of the $\int d(HI)$ curve have been chosen such that the curve coincides with the thermal curve.)

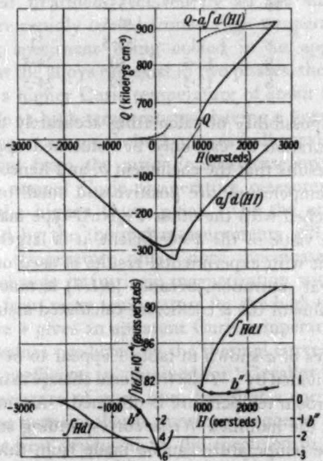


Figure 3. Analysis for Alcomax II (b'' estimated as described in text).

In order to complete the analysis, the coefficients a and b must be found. Consider first the coefficient b associated with the rotational processes. During the initial reduction in field from its maximum value Alcomax II is saturated, and the variation of I with field is small or zero, so that $\int H\,dI$ is also small or zero, as shown in figure 3. Since I is obtained by subtracting two large quantities, $B-H$, small experimental errors will be accentuated, especially in the region of saturation, with the result that whether $\int H\,dI$ is zero or small is not certain.

An approximate analysis of the results can be made by assuming that when the material is saturated the rotational processes involved in magnetization are exactly zero, and therefore, any observed thermal changes are due to changes in intrinsic magnetization. Thermal changes near saturation can then be equated to the first term of the equation and a value of the coefficient a obtained. When a is known the apparent value of b (denoted by b'') can then be found as a function of field, although it must be realized that the values of b'' thus obtained may not be reliable.

In figure 3 the quantity $a\int d(HI)$ and the apparent b coefficient b'' are plotted against the field over the relevant regions of the hysteresis cycle. For negative fields from saturation to about -500 oersteds the $a\int d(HI)$ curve and the thermal curve exactly coincide, and hence b'' is zero over this region. The variation of b'' with field over the rest of the cycle is typical of the results obtained in previous work for regions of the cycle where irreversible changes are known to take place, and the shape of the b'' curve may therefore be primarily due to irreversible processes with the rotation contributions small.

Table 2. Nominal Percentage by weight of the Fe_2NiAl-Type Alloys

	Percentage by Weight					
Material	Al	Ni	Fe	Co	Cu	Nb
Alnico	10	17	55	12	6	—
Alni	13	24	59·5	—	3·5	—
Alcomax II	8	11	55·5	21	4·5	—
Alcomax III	8	13·5	50	24·5	3	1
Alcomax IV	8	13·5	49	24·5	3	2

Owing to the impossibility of calculating accurately the variation of $\int H\,dI$ in the region of saturation b'' can only be estimated approximately. It thus appears from these results that the coefficient b, and hence the fractional change in anisotropy with temperature, are positive and small or even zero. Similar conclusions were reached with the other Fe_2NiAl-type materials investigated.

In estimating the value of the a coefficient, it is largely immaterial whether the method of best fit with experimental results is used or that of assuming no rotation processes near saturation, because $\int d(HI)$ is much greater than $\int H\,dI$. Table 1 shows the value of the a coefficient calculated assuming $\int H\,dI$ is exactly zero at saturation.

A priori, the values of a shown in table 1 appear to be larger than expected. If the coefficient a is divided by $-T$ the fractional change in intrinsic magnetization with temperature at room temperature is obtained. Assuming that the material approximately obeys the modified Weiss corresponding states curve ($J=\frac{1}{2}$), an estimate of the Curie temperature can be made from this value of the rate of change of intrinsic magnetization with temperature, and can be compared with the experimental Curie temperature. A value of *a* of 0·1 corresponds to a Curie temperature of approximately 400°c, whereas the published values of the Curie temperature of the Fe_2NiAl material considered here lie between 730 and 860°c.

An explanation of this large discrepancy between the predicted and observed Curie temperatures is found on examining some of the results obtained by Sucksmith (1939) for the corresponding states curves of the Fe–Al–Ni alloy system; in particular, those for the Fe_2NiAl alloys which have a very similar composition to Alni. The atomic composition of Alni is 52 Fe, 21 Ni, 24 Al and 3 Cu. Figure 4 is based on Sucksmith's work and shows the corresponding states curves for three different heat treatments. Quenching from 800°c is similar to the industrial heat treatment received by all the alloys considered here. Commercial alloys are rapidly cooled from a high temperature and annealed at 500°c, the Alcomax specimens being cooled in an applied magnetic field. Figure 4 indicates that the alloys can exist in two phases, the lower phase obtained by annealing having a higher Curie temperature of about 800°c, and the higher phase (which is stable at high temperatures) having a lower Curie temperature. In the permanent-magnet materials both phases will coexist, their compositions being slightly different from the pure Fe_2NiAl case owing to their different compositions, and a normal Curie temperature measurement will correspond to the phase with the higher Curie point. On the other hand, the Curie temperature indicated by the thermal measurements will be largely associated with the low Curie temperature phase since this phase will give the larger contribution to the change in intrinsic magnetization with temperature. The simple extrapolation from room temperature for the high temperature quenched curve shown in figure 4 gives an apparent Curie temperature of 430°c in rough agreement with the values predicted by the thermal curves.

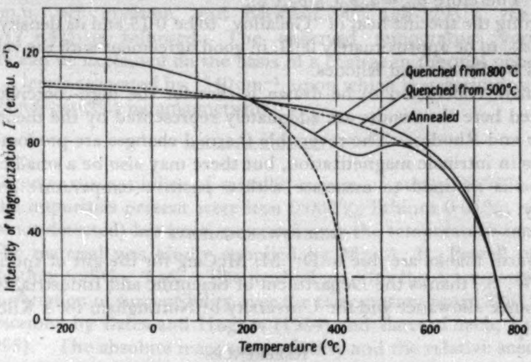


Figure 4. Corresponding states curves of Fe_2NiAl (after Sucksmith).

The value of the coefficient a_G obtained for 'Gecalloy' should be consistent with the Curie temperature of pure iron, but first a correction must be made for the fraction of the material that is non-ferromagnetic. For, experimentally, a_G is the ratio of the heat produced by the specimen to the change in $\int d(HI)$ associated with that heat change. The quantity of heat measured would produce, if there were no plastic binder, a greater temperature change than that recorded, and therefore to obtain the heat change for iron alone one should multiply the observed change by the ratio of the thermal capacity of the material itself, $M_G C_G$, to that of the pure iron, $M_{Fe} C_{Fe}$. As the material is saturated, the change in internal field for a particular particle will be identical with the change in external field, and therefore only the value of I need be corrected in calculating initial values of $\int d(HI)$ for iron alone. This correction is necessary because the true intensity of magnetization of iron is greater than that of the body as a whole in the ratio of the volume of the specimen M_G/ρ_G to that of the iron it contains M_{Fe}/ρ_{Fe}. Therefore $a_{Fe} = a_G \rho_G C_G / \rho_{Fe} C_{Fe}$.

Assuming the specific heat of ' Gecalloy ' to be 0·15 and its density 1·8 g cm^{-3}, one finds a_{Fe} to be approximately 0·03, in good agreement with the value of 0·034 calculated by Stoner and Rhodes.

The main conclusion to be drawn is that for the high-coercivity materials investigated here the results are adequately represented by the theory proposed by Stoner and Rhodes. The reversible thermal changes are predominantly due to changes in intrinsic magnetization, but there may also be a small contribution which one is inclined to associate with a positive temperature coefficient of anisotropy.

ACKNOWLEDGMENTS

Our warm thanks are due to Dr. M. McCaig for the gift of specimens; one of us (A. W. S.) thanks the Department of Scientific and Industrial Research for a maintenance allowance and the University of Nottingham for a Kilby Research Scholarship.

REFERENCES

BATES, L. F., and MARSHALL, G., 1953, *Rev. Mod. Phys.*, **25,** 17.
BATES, L. F., and SHERRY, N. P. R., 1955, *Proc. Phys. Soc.* B, **68,** 304.
DOUCET, V., 1951, *J. Phys. Radium*, **12,** 827.
JONES, R. V., 1951, *Proc. Phys. Soc.* B, **64,** 469.
OKAMURA, T., 1936, *Sci. Reports Tohoku Imp. Univ.*, **24,** 745.
STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, **40,** 481.
STONER, E. C., and WOHLFARTH, E. P., 1948, *Trans. Roy. Soc.* A, **240,** 599.
SUCKSMITH, W., 1939, *Proc. Roy. Soc.* A, **171,** 525.

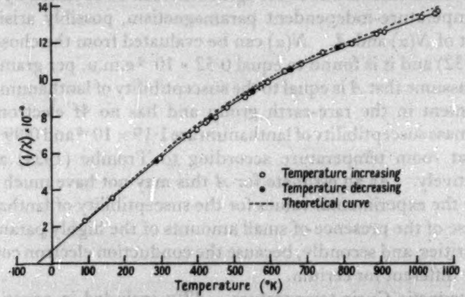
The Magnetic Susceptibility of Metallic Cerium

By L. F. BATES, S. J. LEACH and R. G. LOASBY
University of Nottingham

MS. received 7th July 1955 *and in final form* 4th *August* 1955

Abstract. The magnetic susceptibility of pure metallic cerium has been measured over the temperature range 90° to 1057°K. At 293°K the susceptibility is $17{\cdot}5 \pm 0{\cdot}2$ e.m.u. per gramme, and between 90° and 295°K the relation $\chi = 0{\cdot}00588/(T+42)$ is followed. The observed temperature variation of susceptibility can be explained on the basis of a change in thermal population of ${}^2F_{5/2}$ and ${}^2F_{7/2}$ levels separated by 1340 cm^{-1}, upon which is superimposed a small temperature-independent paramagnetism.

The magnetic susceptibility of metallic cerium of purity 99·94% (the only metallic impurities present were iron 0·002%, lithium 0·05%; other rare earths not detected) has been measured over the temperature range 90° to 1057°K. The material was kindly supplied by Mr. A. R. Powell of Messrs. Johnson, Matthey and Co., Ltd. The method used for the determination of the temperature variation of susceptibility over the temperature range 288° to 1057°K was that described by Bates and Hughes (1954) and Bates, Leach, Loasby and Stevens (1955). The absolute mass susceptibility and the relative susceptibility over the temperature range 90° to 295°K were measured with an apparatus similar to that of Bates and Mallard (1950) using an aqueous solution of nickel chloride as a susceptibility standard (Nettleton and Sugden 1939). The value found at 293°K was $17{\cdot}5 \pm 0{\cdot}2 \times 10^{-6}$ e.m.u. per gramme, and no field variation was detected with fields up to 4500 Oe. The variation of susceptibility is given in the figure which shows the reciprocal mass susceptibility plotted against the absolute temperature. Continuous observations were made over a period of several days and the measurements recorded with increasing and decreasing specimen temperature are indicated in the figure.



Temperature variation of mass susceptibility of metallic cerium.

It is known that there are two structural transformations in the above temperature range, namely at 977°K according to Spedding and Daane (1954) and at 423°K according to Trombe and Föex (1944). From the continuous and reversible results shown in the figure it follows that the changes in susceptibility associated with the transformations are extremely small, i.e. less than 1%.

It is generally accepted that three electrons of cerium are in a conduction band, arising from the 5d and 6s levels of the free atom, and that there is one electron in the 4f subshell. The 4f electron has a $^2F_{5/2}$ ground state which has a magnetic moment of 2·54 Bohr magnetons. The atomic moment calculated from the Curie–Weiss law representing the experimental results over the temperature range 90° to 293°K, $\chi = 0{\cdot}00588/(T+42)$, is 2·58 Bohr magnetons. Now, Van Vleck (1932) has shown that it is important to consider the population of higher levels when $\boldsymbol{k}T$ becomes of the order of the level spacing. Cerium has only two levels arising from spin–orbit coupling, $^2F_{5/2}$ and $^2F_{7/2}$, and if their separation is Δ and the paramagnetic Curie temperature is taken to be -42°K, then the temperature variation of mass susceptibility arising from the population of these levels is given by

$$\chi = \frac{0{\cdot}0008858}{T+42}\left[\frac{38{\cdot}57+164{\cdot}7\exp(-\Delta/\boldsymbol{k}T)}{6+8\exp(-\Delta/\boldsymbol{k}T)}\right]+N(\alpha)+A \qquad \ldots\ldots(1)$$

where $N(\alpha)$ is the temperature-independent paramagnetism which was shown to be present by a more rigorous treatment (Van Vleck 1932), and which will later be shown to be small in the case of cerium; A is the resultant contribution of the diamagnetic core and the conduction electrons and may be positive or negative. If A and $N(\alpha)$ are for the moment neglected, then the temperature dependence of paramagnetism can be fitted by the choice of a suitable value for Δ. For a free cerium ion the separation of the $^2F_{5/2}$ and $^2F_{7/2}$ levels has been reported to be 6060 cm^{-1} (Landolt–Börnstein 1955) and 2240 cm^{-1} (Elliott and Stevens 1953), but it may well be reduced in the metallic state. If it is chosen to be 1340 cm^{-1} the broken line shown in the figure is obtained. This agrees very well with the experimental measurements but closer agreement can be obtained by the inclusion of a small temperature-independent paramagnetism, possibly arising from the resultant effect of $N(\alpha)$ and A. $N(\alpha)$ can be evaluated from the chosen value of Δ (Van Vleck 1932) and it is found to equal $0{\cdot}52 \times 10^{-6}$ e.m.u. per gramme. It may be possible to assume that A is equal to the susceptibility of lanthanum, which is the preceding element in the rare-earth group and has no 4f electron. Reported values for the mass susceptibility of lanthanum are $1{\cdot}19 \times 10^{-6}$ and $0{\cdot}99 \times 10^{-6}$ e.m.u. per gramme at room temperature according to Trombe (1934) and Bommer (1939), respectively. As an estimate for A this may not have much significance, firstly because the experimental values for the susceptibility of lanthanum may be in error because of the presence of small amounts of the highly paramagnetic rare earths as impurities, and secondly, because the conduction electron contribution to A may be very different for cerium.

The paramagnetic Curie temperature -42°K included in equation (1) could arise either from a splitting of the ground state by the crystal field or from a negative exchange interaction. If the splitting of the ground state is of the order of $\boldsymbol{k}T$ then it has been shown (Penney and Schlapp 1932) that a polycrystalline rare earth can follow a Curie–Weiss type of temperature dependence over several hundred degrees. For this to be the origin of the Curie temperature, the splitting would have to be of the order of magnitude 200 cm^{-1}, and this would seem to be surprisingly large for the metallic state. If the excited state is split by the crystal field then equation (1) would not be invalidated since the probability of occupation of the split levels will be almost equal if the splitting is small compared to Δ.

If the Curie temperature arises from an exchange interaction then this is probably of the Zener type, i.e. involving coupling with the conduction electrons (Bagguley and Stevens, unpublished) and may therefore give rise to co-operative phenomena at low temperatures. The behaviour at low temperatures may however be complicated if crystal field splittings are also present.

It is therefore concluded that the temperature dependence of susceptibility of metallic cerium from 90°K to its melting point can be satisfactorily explained by the thermal population of the $^2F_{5/2}$ and $^2F_{7/2}$ levels separated by 1340 cm^{-1} and a small superimposed temperature independent paramagnetism.

ACKNOWLEDGMENT

We are indebted to Dr. K. W. H. Stevens for many helpful discussions.

REFERENCES

BATES, L. F., and HUGHES, D., 1954, *Proc. Phys. Soc.* B, **67,** 28.

BATES, L. F., LEACH, S. J., LOASBY, R. G., and STEVENS, K. W. H., 1955, *Proc. Phys. Soc.* B, **68,** 181.

BATES, L. F., and MALLARD, J. R., 1950, *Proc. Phys. Soc.* B, **63,** 520.

BOMMER, H., 1939, *Elektrochem.*, **45,** 357.

ELLIOTT, J., and STEVENS, K. W. H., 1953, *Proc. Roy. Soc.* A, **218,** 553.

LANDOLT–BÖRNSTEIN, 1955, *Atom und Molekularphysik* (Berlin : Springer).

NETTLETON, H. R., and SUGDEN, S., 1939, *Proc. Roy. Soc.* A, **173,** 313.

PENNEY, W. G., and SCHLAPP, R., 1932, *Phys. Rev.*, **41,** 194.

SPEDDING, F. H., and DAANE, A. H., 1954, *J. Metals, N.Y.*, 504.

TROMBE, F., 1934, *C. R. Acad. Sci., Paris*, **201,** 652.

TROMBE, F., and FÖEX, M., 1944, *Ann. Chim.*, **19,** 417.

VAN VLECK, J. H., 1932, *The Theory of Electric and Magnetic Susceptibilities* (Oxford : University Press).

A Photochemical Reaction in the Human Retina

By G. S. BRINDLEY
Physiological Laboratory, University of Cambridge

MS. received 4*th April* 1955, *and in final form* 31*st May* 1955

Abstract. The way in which the disturbance of a colour match by adaptation is affected by the brightness of the adapting field is investigated. It is inferred from the results that the disturbance is due to a photochemical reaction in a pigment for which the apparent value of the product of quantum efficiency and molar decadic extinction coefficient at $\lambda\ 0{\cdot}574\,\mu$ is about 420 000. Reasons are given for believing that the true value is probably lower. Three hypotheses concerning the nature of the pigment are discussed, and it is concluded that the most likely of them is that it is the receptor pigment of the 'red-receptive' mechanism of trichromatic theory. If so, this must have the very high optical density in the receptors of about 0·98.

§ 1. INTRODUCTION

WRIGHT (1936) found that, though colour matches made by the normal human eye are remarkably stable to moderate degrees of adaptation to white or coloured light, with very bright adapting lights they may be disturbed. Detailed investigation of this disturbance (Brindley 1953) revealed:

(i) that the factor by which a given adaptation alters the ratio of amounts of fixed red and green primaries required to match test wavelengths between 0·48 and $0{\cdot}62\,\mu$ differs for different test wavelengths;

(ii) that the manner in which a match of yellow against red + green is disturbed by adaptation is independent of the adapting wavelength, and that the amounts of light at different wavelengths required to produce equal *magnitudes* of disturbance define an action spectrum which agrees within experimental error with a likely spectral sensitivity curve for the 'red' mechanism of trichromatic vision;

(iii) that the disturbance is closely similar to a disturbance of colour matches found when all the light making up both photometric fields is passed through the periphery instead of the centre of the pupil;

(iv) that all the disturbances, both with adaptation and with decentring, are within experimental error of those implied by the hypothesis that there are three classes of receptors active in foveal photopic vision, and that of these the 'red' receptors detect light by means of a photolabile pigment present in high peak optical density, which is reduced by 0·5 by the adaptations used in the experiments and by 0·2 by the oblique incidence.

It was inferred from (i) that the disturbance was not due to production or destruction of a screening pigment in front of all the receptors and from (ii) that a single photochemical reaction was involved; and it was suggested, on the evidence of (iii) and (iv) together, that the hypothesis set out in (iv) might be correct.

The present experiments and arguments provide further evidence related to these inferences.

§ 2. THEORY

If the disturbance of the colour match is due to bleaching of the red-receptive pigment, which in the unadapted eye has a peak density high enough to cause substantial self-screening for light of wavelengths near that of maximum absorption, then we can deduce the relations between the disturbance of the colour match and the concentration of pigment remaining unbleached, and between the concentration of pigment remaining, the amount of bleaching light, and the product of extinction coefficient and quantum efficiency of bleaching. The relation between the disturbance of the colour match and the amount of bleaching light can be tested experimentally, and if the experimentally determined relation agrees with that predicted, the product of extinction coefficient and quantum efficiency of bleaching can be calculated.

2.1. *Relation of Concentration of Pigment remaining to Amount of Bleaching Light*

The following symbols are used: q is the number of quanta per cm^2 of retina in the bleaching light, c_0, c the number of red-receptive pigment molecules per cm^2 of cross section of the receptors before and after bleaching respectively, α_b the molecular extinction coefficient of the red-receptive pigment for the bleaching wavelength in Napierian units, γ the quantum efficiency of bleaching.

The optical density of the receptors for the bleaching wavelength is $\alpha_b c \log e$, the fraction of incident light absorbed is $1-\exp(-\alpha_b c)$, and

$$dc/dq = -\gamma\{1-\exp(-\alpha_b c)\},$$

whence

$$q = -\frac{1}{\gamma}\left[c + \frac{1}{\alpha_b}\ln\{1-\exp(-\alpha_b c)\}\right] + \text{const.}$$

Determining the constant of integration from the relation that $c = c_0$ when $q = 0$, we find

$$q = -\frac{1}{\alpha_b\gamma}[\alpha_b c - \alpha_b c_0 + \ln\{1-\exp(-\alpha_b c)\} - \ln\{1-\exp(-\alpha_b c_0)\}]. \quad \ldots\ldots(1)$$

When $\alpha_b c_0$ is small, this reduces to the familiar form

$$c = c_0 \exp(-\alpha_b \gamma q). \quad \ldots\ldots(2)$$

2.2. *Relation of $\Delta \log R$ to the Concentration of Pigment remaining unbleached*

The symbols used are: R, G, Y are the amounts of red, green and yellow test lights in a colour match for the unadapted eye, R', G' the corresponding quantities of red and green for the adapted eye. A_r, A_g, A_y are the sensitivities of the 'red-receptive' mechanism to the red, green and yellow test wavelengths in the unadapted eye, a_r, a_g, a_y the corresponding quantities of the adapted eye. B_r, B_g and B_y are the corresponding sensitivities of the 'green-receptive' mechanism (the shape of whose spectral sensitivity curve is supposed not to be affected by the adaptation). $n = A_r B_g / B_r A_g$. We have

$$RA_r + GA_g = YA_y \qquad R'a_r + G'a_g = Ya_y$$

$$RB_r + GB_g = YB_y \qquad R'B_r + G'B_g = YB_y$$

whence by eliminating G, G' and Y

$$\Delta \log R = \log R' - \log R = -\log \frac{(a_r B_g - B_r a_g)(A_y B_g - B_y A_g)}{(A_r B_g - B_r A_g)(a_y B_g - B_y a_g)}.$$

If the sensitivities A and a depend on a photochemical reaction in the pigment bleached by the adapting light, then, for any wavelength λ,

$$\frac{a}{A}=\frac{1-\exp(-\alpha_\lambda c)}{1-\exp(-\alpha_\lambda c_0)}$$ where α_λ is the Napierian molecular extinction coefficient at wavelength λ,

$\simeq c/c_0$ for those wavelengths for which $\alpha_\lambda c_0$ is very small.

For the experimental test, the wavelengths 0·68, 0·57 and 0·55 μ were chosen, with the aim that, if the spectral sensitivity curve of the 'red-receptive' mechanism were represented by the luminosity curve of the eye in the monochromacy following violet and blue-green adaptation (Brindley 1953) or by the luminosity curve of deuteranopes on a 2° (Pitt 1935) or 12′ (Willmer 1949) field, or to a tolerable approximation if it were that given by Pitt (1944), the densities $\alpha_\lambda c_0 \log e$ would be for the yellow and green test lights nearly equal to the maximum density, and for the red test light very small.

Then $$\frac{a_g}{A_g}=\frac{a_y}{A_y}=\frac{1-\exp(-\alpha_m c)}{1-\exp(-\alpha_m c_0)}$$

where α_m is the maximum molecular extinction coefficient, and $a_r/A_r=c/c_0$.

Besides simplifying the equations, this choice of test wavelengths has the advantage, both predicted from the present theory, assuming the sensitivity curves of Pitt (1944) or of Brindley (1953) for the red-receptive and green-receptive mechanisms, and verified in the present experiments, that $\Delta\log G$ is for all degrees of adaptation so small as to be very nearly undetectable; and in the principal experiments the brightness of the green-matching light was set for the unadapted eye and then left unchanged throughout every adaptation and recovery. The experiments were thereby certainly made easier, and very probably more accurate.

For the chosen test wavelengths,

$$\Delta\log R=-\log\frac{(a_rB_g-B_ra_g)(A_yB_g-B_yA_g)}{(A_rB_g-B_rA_g)(a_yB_g-B_ya_g)}$$
$$=-\log\frac{n(c/c_0)\{1-\exp(-\alpha c_0)\}-\{1-\exp(-\alpha c)\}}{(n-1)\{1-\exp(-\alpha c)\}}\quad\ldots\ldots(3)$$

When αc_0 is small, so that the value of c can be taken from equation (2) and terms of the third and higher orders be neglected, equation (3) approximates to

$$\Delta\log R=-\frac{n\alpha(c_0-c)}{2(n-1)}\log e$$
$$=-\{1-\exp(-\alpha_b\gamma q)\}\frac{n\alpha c_0}{2(n-1)}\log e.\quad\ldots\ldots(4)$$

§ 3. EXPERIMENTAL METHODS

The apparatus and methods of calibration were those described in Brindley (1953), and the sizes of adapting field (3°) and test fields (1°) were the same. An artificial pupil of 2 mm diameter was used. It was verified that the adapting light never caused the subject's pupil to constrict to less than this size, so that the retinal illumination was always that calculated from the artificial pupil. For the test fields the illumination was 30 trolands; for the adapting field it varied as described later.

The results presented in detail were obtained with the author as subject. Very similar results obtained with another subject are briefly described.

§ 4. RESULTS

Figure 1 shows the course of recovery towards the normal (unadapted) colour match after adaptation for 10 seconds at various retinal illuminations to light of modal wavelength 0·574 μ and band-width at half maximum 0·058 μ, determined by a dyed gelatine filter (Ilford No. 626). The curves have been extrapolated back to the instant of removing the adapting light, and in figure 2 the intercepts are plotted against the number of quanta falling on each square centimetre of

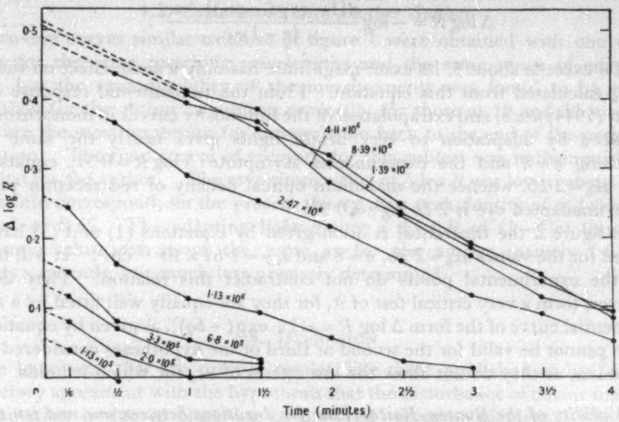


Figure 1. Time-course of recovery of a colour match disturbed by adaptation to different amounts of yellow light (values in trolands × seconds).

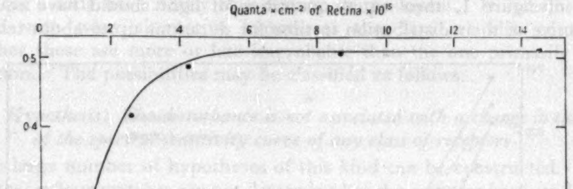


Figure 2. Amount of disturbance of a colour match, extrapolated to zero time, as a function of the amount of adapting light.

retina during the period of adaptation, assuming that all the adapting light was of wavelength $0{\cdot}574\,\mu$ and that one troland at this wavelength gives $1{\cdot}036 \times 10^9$ quanta $\mathrm{sec}^{-1}\,\mathrm{cm}^{-2}$ of retina. This estimate is based on a posterior nodal distance of 16·8 mm, Ludvigh and McCarthy's (1938) value of 59·4% for the transmission of the eye at $0{\cdot}580\,\mu$ and a luminosity of 598 lumens per watt at $0{\cdot}574\,\mu$.

The experimental points seem to approach an asymptote at about $\Delta \log R = 0{\cdot}51$ for high retinal illuminations. Now the limit of equation (3) as c tends to zero is

$$\Delta \log R = -\log \frac{n\{1-\exp(-\alpha c_0)\} - \alpha c_0}{(n-1)\alpha c_0}. \qquad \ldots\ldots(5)$$

If n exceeds about 5, its exact magnitude has only a small effect on the value of αc_0 calculated from this equation. From the fundamental response curves of Pitt (1944) $n \simeq 8$, and extrapolation of the luminosity curves in monochromacies produced by adaptation to very bright lights gives nearly the same value. Assuming $n = 8$ and the experimental asymptote $\Delta \log R = 0{\cdot}51$, equation (5) gives $\alpha c_0 = 2{\cdot}26$, whence the maximum optical density of red receptor pigment in the unadapted eye is $2{\cdot}26 \log \mathrm{e} = 0{\cdot}98$.

In figure 2 the theoretical relation given by equations (1) and (3) has been plotted for the values $\alpha c_0 = 2{\cdot}26$, $n = 8$ and $\alpha_b \gamma = 1{\cdot}61 \times 10^{-15}\,\mathrm{cm}^2$. It will be seen that the experimental points do not contradict this relation. They do not, however, form a very critical test of it, for they are equally well fitted by a simple exponential curve of the form $\Delta \log R = a\{1 - \exp(-bq)\}$, as given by equation (4), which cannot be valid for the second or third of the hypotheses considered in the discussion, as they do not allow the low values of αc_0 for which it holds.

4.1. *Validity of the Bunsen–Roscoe Law for durations between one and ten seconds*

If the disturbance of colour matches is due directly to a photochemical reaction in a pigment whose regeneration follows the time course of the recovery curves of figure 1, then equal quantities of light should have equal effects irrespective of their distribution in time for durations up to about ten seconds.

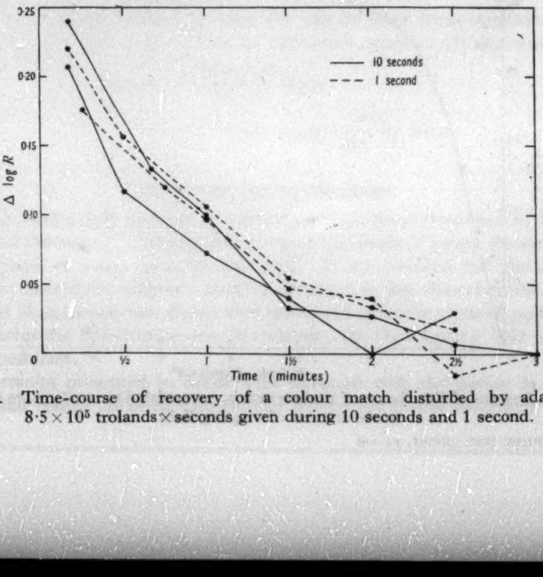


Figure 3. Time-course of recovery of a colour match disturbed by adaptation to $8{\cdot}5 \times 10^5$ trolands × seconds given during 10 seconds and 1 second.

If it is due to a neural mechanism, there is no necessary reason why this relation should hold for durations above the summation time for equal sensation from two flashes, which at high luminance is about 0·04 second (Brindley, unpublished observations). Figure 3 shows the course of recovery of a colour match disturbed by adaptation to $8{\cdot}5 \times 10^6$ trolands × seconds, in two experiments spread over ten seconds, and in two experiments concentrated within one second. No departure from the Bunsen–Roscoe law is detected.

4.2. *Experiments with another Subject*

Recovery curves similar to those of figure 1 were obtained with one other subject for the same matching wavelengths and the same range of adapting lights; but the reproducibility of the measurements was found to be much worse than for the author as subject, especially for those at 10 and 30 seconds, which are the most important for extrapolation back to the end of the period of adaptation. The time course of recovery for this subject was indistinguishable from that for the author. The asymptotic value of $\Delta \log R$ was lower, about 0·35. This would correspond, on the present theory, to a peak density of red receptor pigment of 0·65. The adapting light giving a $\Delta \log R$ of $(1-1/e)$ times the asymptotic value was about the same as for the author, namely $1{\cdot}6 \times 10^6$ trolands × seconds, but much less precisely determined.

§ 5. DISCUSSION

We have seen that the present results, like those published earlier, are in satisfactory agreement with the hypothesis that the disturbance of colour matches by adaptation is due to bleaching of the 'red' receptor pigment. The peak density of this pigment which they require, namely 0·98, is surprisingly high, but perhaps not very unlikely, in view of Denton and Wyllie's (1955) finding that the frog's rods have nearly this density of rhodopsin. It remains to be considered with what alternative hypotheses the data are also consistent, and whether these are more or less improbable than the one primarily under consideration. The possibilities may be classified as follows.

5.1. *First Hypothesis: The disturbance is not associated with a change in the shape of the spectral sensitivity curve of any class of receptors*

A very large number of hypotheses of this kind can be constructed. They all imply that colour matches are not determined at the receptor level, and hence that there are both more than three kinds of light-sensitive pigment and more than three classes of receptors active in foveal photopic vision. They fail, unless additional and unlikely *ad hoc* hypotheses are made, to explain Grassmann's laws of additivity of colour mixture, which follow automatically from the hypothesis that there are three kinds of pigment or three classes of receptors or both; they fail, unless further unlikely *ad hoc* hypotheses are made, to explain the stability of colour matches to moderate degrees of adaptation to coloured lights, which also follows automatically if there are three kinds of pigment or three classes of receptors; and they fail unless yet more *ad hoc* hypotheses are made, to explain any of the quantitative aspects of the disturbance by adaptation to very bright lights, all of which are adequately explained by the hypothesis of bleaching of the 'red' receptor pigment.

So complex do hypotheses of this first kind become when modified to fit all the experimental facts that it seems reasonable to neglect them while there are simple hypotheses that fits the facts as well.

If the spectral sensitivity curves of receptors are altered by adaptation, the alteration must be due to changes in the amounts of receptive pigments or of screening pigments and much the simplest supposition is that the changes are directly due to the absorption of light by these pigments themselves, though it is difficult to exclude altogether the possibility that they may be secondary to the absorption of light by other substances.

The fact that the disturbance obtained with any monochromatic adapting light can be reproduced with any other monochromatic light, if the brightness is adjusted, proves that the absorption of light by only one substance is concerned, or if by more than one, that these have nearly the same absorption spectrum, or else are not themselves screening pigments, but affect secondarily the same screening or receptive pigment.

5.2. *Second Hypothesis: A non-receptor pigment, acting as a filter in front of the receptors, is bleached or photosynthesized*

This cannot be a general screening pigment in front of all the receptors equally (Wright 1936, Brindley 1953). It must lie in front of the receptors individually, screening one class and not another. Such pre-receptoral colour filters, though they occur in reptiles and birds, have never been observed in mammalian retinae. If such a pigment is *photosynthesized*, then to explain the results of figure 2 of Brindley (1953) it must have a low density in the green and a very high density in the red, nearly constant over the range 0·62–0·70 μ, and this is very unlikely. If it is *bleached*, it requires a more plausible absorption spectrum.

5.3. *Third Hypothesis: Only two receptor pigments are substantially active in the dichromatic red–yellow–green range of the spectrum, and one of them has a high density, which is diminished by adaptation*

If so, then on the evidence already presented (Brindley 1953) this must be the 'red-receptive' pigment, which must have the very high peak density of 0·98 if, as the present calculations assume, the product of bleaching is colourless, and cannot be much less dense if the product of bleaching is coloured.

5.4. *Fourth Hypothesis: There are more than two receptor pigments active to a substantial degree in the dichromatic red–yellow–green range of the spectrum, the spectral range being dichromatic because these pigments are contained in only two classes of receptors*

If so, all the experimental results on the disturbance of colour matches by *adaptation* can be explained if one of the pigments is much more photosensitive than the others, so that for all adapting wavelengths it is preferentially bleached; but the agreement of the direction and form of the disturbance with that calculated on the third hypothesis becomes a mere coincidence, and the very similar disturbance of colour matches on passing the light through the periphery of the pupil requires a separate and, it seems, unrelated explanation.

Of these four hypotheses, none has been disproved, but the first may fairly be discarded on the grounds of its extreme complexity when modified to fit all the facts. Of the other three, that of bleaching of the red-receptive pigment is the simplest and explains quantitatively the greatest number of experimental observations. In the absence of any conclusive evidence it may for this reason be regaded as the most likely.

5.5. *Photosensitivity of the Pigment Responsible*

For the theoretical relation of equations (1) and (3) to fit the data of figure 2, $\alpha\gamma$ must have the value $1{\cdot}61 \times 10^{-15}\,\mathrm{cm}^2$ at $\lambda\ 0{\cdot}574\ \mu$. If the second or fourth of the above hypotheses were correct, instead of the third, as assumed in equations (1) and (3), it could be a little lower; for example, if the value of $\Delta \log R$ were proportional to the amount of pigment bleached, and the relevant pigment were scarce enough for equation (2) to hold, the best-fitting value of $\alpha\gamma$ would be about $6 \times 10^{-16}\,\mathrm{cm}^2$.

In decadic molar units, the former estimate corresponds to $\epsilon\gamma = 420\,000$. This is eighteen times the photosensitivity of visual purple (Schneider, Goodeve and Lythgoe 1939), and larger than the largest value of ϵ quoted by Braude (1950) for polyenes and aromatic hydrocarbons. If the quantum efficiency γ is unity it implies on Braude's theory a much larger effective chromophore area than that of visual purple. This somewhat unlikely conclusion would be avoided if either the quantum efficiency were greater than unity or an optical system within the retina concentrated the light on the pigment, making the apparent photosensitivity higher than it would be in uniform illumination. The first of these possibilities seems for biological reasons a little unlikely; the retina must contain an amplifier, and it is not unlikely that the first stage of this is a chain reaction within the receptors, but it would appear uneconomical for such a chain reaction to use the receptive pigment itself. The second possibility, on the other hand, might have been expected from the directional sensitivity of the retina even if the high apparent photosensitivity of the red-receptive pigment had not been found. A device causing the sensitivity of a receptor to light from one direction to be higher than to light from another might work merely by throwing away sensitivity in the unfavoured direction, or by using it to gain sensitivity in the favoured. Orientation of the chromophore groups perpendicularly to the favoured direction of incidence (Schmidt 1938, Denton 1954) is a mechanism of the latter kind, but gaining only a factor of 1·5; another is the hypothesis of O'Brien (1946) of multiple total internal reflection in a tapering tube; and a third is the hypothesis that an intra-retinal lens, perhaps in the inner segment of each cone, concentrates the light on the pigment in the outer segment. Some such device would be more economical than one which merely threw away sensitivity in the unfavoured direction, and might for this reason be regarded as more likely.

5.6. *Objective Demonstration of Bleaching of a Foveal Pigment*

If a retinal pigment present in high density is bleached, it should be possible to detect changes after adaptation in the amount of light of different wavelengths reflected from the fovea. On the present theory these changes need be by no means so large as the hundredfold that would be expected if all the light reflected had passed twice through receptors of density 0·98, since red receptors certainly do not occupy the whole foveal area and, if they are very dense, will contribute much less light than in proportion to the area which they occupy. Using a modification of the apparatus described by Rushton (1952), such changes in reflecting properties have been detected (Brindley and Rushton 1955), and are being investigated in detail (Rushton, in course of publication). The maximum change in reflection fraction on bleaching is about 0·12 log unit.

REFERENCES

Braude, E. A., 1950, *J. Chem. Soc.*, **1950**, 379.
Brindley, G. S., 1953, *J. Physiol.*, **122**, 332.
Brindley, G. S., and **Rushton**, W. A. H., 1955, *J. Physiol.*, **128**, 59P
Denton, E. J., 1954, *J. Physiol.*, **124**, 16P.
Denton, E. J., and **Wyllie**, J. H., 1955, *J. Physiol.*, **127**, 81.
Ludvigh, E., and **McCarthy**, E. F., 1938, *Arch. Ophthal.*, *N.Y.*, **20**, 37.
O'Brien, B., 1946, *J. Opt. Soc. Amer.*, **36**, 506.
Pitt, F. H. G., 1935, *Med. Res. Coun.* Spec. Rep. No. 200; 1944, *Proc. Roy. Soc.* B, **132**, 101.
Rushton, W. A. H., 1952, *J. Physiol.*, **117**, 47P.
Schmidt, W. J., 1938, *Kolloidzschr.*, **85**, 137.
Schneider, E., **Goodeve**, C., and **Lythgoe**, R., 1939, *Proc. Roy. Soc.* A, **170**, 102.
Willmer, E. N., 1949, *J. Physiol.*, **110**, 422.
Wright, W. D., 1936, *J. Physiol.*, **87**, 23.

An Interferometer for measuring the Spatial Frequency Response of a Lens System

By L. R. BAKER
Technical Optics Section, Imperial College, London

MS. received 1st *June* 1955 *and in final form* 20*th August* 1955

Abstract. The new approach to optical image formation from the standpoint of Fourier methods has led to the need for characterizing a lens by its response to spatial frequencies. Methods of measuring this reponse by means of photometric image scanning are neither simple nor accurate, especially when the object is illuminated with partially coherent light, whereas the instrument described here gives the frequency response directly and for any specified conditions of illumination. The apparatus is an interferometric photometer consisting basically of a shearing interferometer giving any amount of shear but with no tilt. It is shown that under certain conditions the total light flux in the sheared interferogram is a measure of the frequency response of the lens producing it. Some experimental results are compared with those obtained theoretically by Hopkins for the case of a defocused optical system.

§ 1. INTRODUCTION

THE correspondence between object and image is the principal concern of both lens designers and users. In most cases it is desirable for the image to resemble the object as nearly as possible, although this relationship is sometimes deliberately falsified so as to render a normally unobservable object easily visible, as, for example, in the phase contrast microscope. The mathematical problem of determining directly the distribution of intensity in the image of an extended object under a particular type of illumination is formidable, but with the application of convolution-type theorems and the methods of Fourier analysis a general solution has been found possible.

Essentially, this approach to image formation was adopted both by Abbe, who applied Fourier analysis to a coherently illuminated object, and by Rayleigh, (1896) who studied the image of a self-luminous object and actually gave expressions for the amplitude of the spatial frequencies present. The power of the method, however, was only realized after Duffieux (1946) had given a mathematical treatment of image formation using expressions for the degree of transmission of spatial frequencies from the object to the image plane. Steel (1952, 1953) using Fourier transforms calculated the distribution of intensity in the image of an incoherently illuminated line in the presence of small aberrations, and Hopkins (1951, 1953) showed how Duffieux's treatment of coherent and incoherent image formation could be generalized to cover the case of partial coherence. Mention should also be made of Zernike's (1938) early paper on partial coherence together with a recent contribution by Blanc-Lapierre and Dumontet (1955) who have applied the general theory of random functions to this problem. A new theory of aberration tolerances (Hopkins 1955 a) has been based on the transmission factor or frequency response of a lens, so that the measurement of this characteristic is now probably the most useful single test of lens performance, since it relates the lens both to the design and also to its image forming properties under any chosen conditions.

Whilst it is true that the frequency response of a lens may in principle be determined from the distribution of intensity in the image of a given object one can rarely obtain accurate results in practice particularly if the object is illuminated with partially coherent light. A considerable simplification results from the use of one-dimensional test objects, whilst at the same time there is no loss of generality (see Hopkins 1955 b). The object chosen must be analysed into its Fourier spectrum and then the extent to which this has been transmitted can be found by applying Fourier analysis to the distribution of intensity in the image. A one-dimensional sine grating object has often been considered since, as was shown by Selwyn (1948), the image is then still a sine curve, and it is only necessary to measure its reduction in contrast and lateral displacement. Simpler types of object such as the slit and half plane were used by Lindberg (1954) and others. The frequency response when using a slit object is the Fourier transform of the distribution of intensity in the image, and for a half plane it is the transform of the image intensity gradient.

All these methods depending on photometric image scanning require an accurate knowledge of the type of object and illumination used. If the mathematical transformation is simple as in the case of a sine grating then the object is difficult to realize practically, and conversely if the object is physically simple as for a slit then the mathematical analysis to follow is relatively more complex. The interference photometer to be described gives the frequency response of a lens directly for any conditions of illumination and without the use of a particular type of object. The apparatus is essentially an analogue computer incorporating the lens under test, and the theoretical basis of the instrument will now be given.

§ 2. THE THEORY OF THE INTERFERENCE PHOTOMETER

Using the Fourier approach to optical image formation (see Hopkins 1953) a one-dimensional test object with complex transmission $E(u)$ may be expressed in terms of its equivalent Fourier series as

$$E(u)=\frac{1}{(2\pi)^{1/2}}\sum_{m} a(m)\exp[imu] \qquad \ldots\ldots(1)$$

where $a(m)$ is the amplitude of component m.

If each component is perfectly transmitted then the distribution of intensity in the image $I'(u')$ will be given by

$$I'(u')=|E(u')|^2=\frac{1}{2\pi}\sum_{m}\sum_{p} a(m)a^*(p)\exp[i(m-p)u'] \qquad \ldots\ldots(2)$$

where $a^*(p)$ is the complex conjugate of $a(p)$.

In practice, however, some components will be absent from the image due to the finite aperture of the lens and those that are transmitted may suffer attenuation. It has been shown that under these circumstances the actual image distribution is

$$I'(u')=\frac{1}{2\pi}\sum_{m}\sum_{p} C(m,p)\,a(m)\,a^*(p)\exp[i(m-p)u'] \qquad \ldots\ldots(3)$$

where the complex frequency response is given by

$$C(m,p)=2\pi\iint_{-\infty}^{+\infty}\gamma(x,y)\,f(x+m,y)\,f^*(x+p,y)\,dx\,dy. \qquad \ldots\ldots(4)$$

In this expression $f(x, y)$ is the pupil function of the lens and $\gamma(x, y)$ is the 'effective source'. For coherent illumination $\gamma(x, y)$ is a Dirac delta function, for incoherent illumination it is unity for all values of the variables, and in the case of partial coherence it is unity over a limited region of the range of integration and is zero outside. $C(m, p)$ is in general a complex quantity, whose modulus is the contrast transmission of the term (m, p), and whose argument denotes a lateral shift of this component. If the lens aberrations are symmetrical, as for spherical aberration, the response function is wholly real, but if an asymmetrical aberration such as coma is present its imaginary part must also be found.

The photometric scanning methods previously referred to involve the choice of a suitable object function $E(u)$ and then $C(m, p)$ is found by a measurement of the function $I'(u')$ followed by an appropriate mathematical analysis. The interference photometer gives $C(m, p)$ directly by providing an experimental solution of the integration in equation (4).

A method of studying the frequency response of a lens which is based on a measurement of the total light flux in a sheared interferogram was first suggested by Hopkins (1955 c), and during the past two years an interference photometer employing this principle has been constructed which is not only phase sensitive but should also permit measurements of the frequency response accurate to within 1%.

The interferometric unit is to be seen on the extreme right of the schematic representation of the complete interference photometer shown in figure 1.

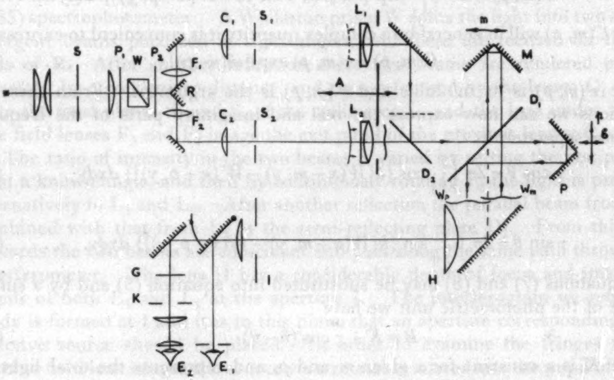


Figure 1. Schematic representation of the complete interference photometer.

It is a simple modification of the Michelson type of interferometer, where the plane mirrors have been replaced by the roof mirrors M and P, in order to provide for a lateral shear of the wave fronts coming from each mirror. S_2 is a monochromatic point or slit source at the focus of the lens under test L_2, giving the aberrant wave front shown. Lateral displacements $m/2$, $p/2$ of the mirrors M and P give a shear m, p to the components W_m, W_p, whose optical path difference is δ.

The complex amplitude of the emergent wave front is given by

$$f(x, y) = \exp [ikW(x, y)]$$

where $W(x, y)$ is the wave aberration and k is the propagation constant $2\pi/\lambda$. The amplitude at some elementary area in the interference pattern produced by the interaction of W_m and W_p will be given by

$$\exp [ikW(x+m, y)] + \exp [ikW(x+p, y)+i\delta]$$

if the wave fronts are taken to have the same amplitude equal to unity. The intensity is the squared modulus of this expression and the total light flux in an area I of the interferogram is therefore given by

$$B = \iint_{\mathrm{I}} |\exp [ikW(x+m, y)] + \exp [ikW(x+p, y)+i\delta]|^2 \, dxdy.$$

When expanded this expression becomes

$$B = 2\iint_{\mathrm{I}} dxdy + 2\cos\delta \iint_{\mathrm{I}} \cos \{k[W(x+m, y) - W(x+p, y)]\} \, dxdy$$
$$+ 2\sin\delta \iint_{\mathrm{I}} \sin \{k[W(x+m, y) - W(x+p, y)]\} \, dxdy. \quad \ldots (5)$$

Returning now to equation (4) we saw that $\gamma(x, y)$ was equivalent to limiting the range of integration, and so we can replace the infinite limits by the range I which is the area common to the two displaced pupils and the effective source. Introducing also the aberration function we may now rewrite the frequency response as

$$C(m, p) = 2\pi \iint_{\mathrm{I}} \exp \{ik[W(x+m, y) - W(x+p, y)]\} \, dxdy.$$

Since $C(m, p)$ will in general be a complex quantity it is convenient to express it as

$$C(m, p) = \tau(m, p) \exp [i\theta(m, p)] \quad \ldots\ldots (6)$$

where $\tau(m, p)$ is its modulus and $\theta(m, p)$ is the argument. From these two equations we can now express the real and imaginary parts of the frequency response separately as

$$\tau \cos\theta = 2\pi \iint_{\mathrm{I}} \cos \{k[W(x+m, y) - W(x+p, y)]\} \, dxdy \quad \ldots (7)$$

and

$$\tau \sin\theta = 2\pi \iint_{\mathrm{I}} \sin \{k[W(x+m, y) - W(x+p, y)]\} \, dxdy. \quad \ldots\ldots (8)$$

Equations (7) and (8) may be substituted into equation (5) and by a suitable choice of the photometric unit we have

$$B = K + \tau \cos(\delta - \theta) \quad \ldots\ldots (9)$$

where K is a constant for a given m and p, and represents the total light flux when no interference takes place.

The modulus of the frequency response is therefore the amplitude of the variation of the total light flux B as the path difference δ changes, and the argument is the phase shift. If the plane wave front from L_1 is used, τ will be equal to K and θ is zero. The argument may thus be measured as the phase difference existing between the flux modulation associated with the aberrant and the plane wave front. It should be noted that if on the other hand a wave front with nominally infinite aberration, or zero frequency response, were used $\tau = 0$ and $B = K$ and in this case the total light flux is completely independent of δ.

If the object is incoherently illuminated the area of integration I will be the area common to the two displaced wave fronts, but in the case of partial coherence, as when testing microscope objectives, for example, it is only necessary to further restrict this area by means of an aperture corresponding in size to the effective source. If the object is coherently illuminated this aperture has a vanishingly small area since in this case the frequency response is given by the pupil function of the lens.

In the experimental results given here the lens tested was a telescope doublet to be used with a self-luminous object and in this case there is no interaction between the m and p terms of equations (2). We may therefore put $m=0$ and $p=s$, where s is the relative shear of the wave fronts. The Fourier variable s may be written in terms of number of lines per unit length R according to $s=\lambda R/\sin\alpha$ where α is the angular inclination to the axis of rays from the pupil $x^2+y^2=1$. When $s=2$ the value of R is the object frequency just on the limit of the resolving power of the lens.

§ 3. THE INTERFERENCE PHOTOMETER

The instrument constitutes essentially a two-beam flicker photometer in which one beam contains the lens under test and the other a good quality collimator. A schematic representation of the complete photometer is shown in figure 1 (Baker and Hopkins 1955).

A sodium lamp on the extreme left is focused on to the source slit S which is itself imaged on to the roof mirror R. The light intensity in the two beams is controlled by the polarization unit P_1WP_2 similar to that employed in the Hardy (1935) spectrophotometer. A Wollaston prism W splits the light into two slightly divergent beams polarized at right angles, and these are focused on the two faces of R. After another reflection these two beams are rendered parallel, separated by a convenient distance, and then refocused by the lenses C_1 and C_2 on to the focal planes S_1 and S_2 of the collimator L_1 and the lens under test L_2. The field lenses F_1 and F_2 image the exit pupil of the previous lens at C_1 and C_2.

The ratio of intensity in the two beams is varied by setting the first polaroid P_1 at a known angle, and then by continuously rotating P_2 the light is presented alternatively to L_1 and L_2. After another reflection the parallel beam from L_1 is combined with that from L_2 at the semi-reflecting plate D_2. From this point onwards the two beams are coincident and pass along the same path through the interferometer. The lens H has a considerable depth of focus and images the pupils of both L_1 and L_2 at the aperture I. The interferogram we require to study is formed at I and it is in this plane that an aperture corresponding to the effective source should be placed. In order to examine the fringes visually another field lens images the source via a rotatable mirror into the pupil of the eye at E_1. For the purpose of a measurement, however, the beam is allowed to pass on and come to a focus at the aperture K which is viewed by E_2 through an eyepiece. A high transmission diffusing element G (Baker 1955) serves to produce a uniform illumination at K.

The source S is a symmetrically opening spectrometer slit and its width is such that the wave fronts are coherent in the direction of shear (see Hopkins 1951). The height of the slit is limited only by off-axis aberrations and vignetting. The polaroid P_1 is mounted in a graduated circle which can be read to 0·01°. A Nicol prism should not be used here as it will almost certainly be mounted slightly off axis in which case rotation will introduce optical path differences between the interfering beams. The speed of rotation of P_2 is about 5 c/s for visual observation, but this can be increased to 60 c/s should an electronic detector be employed. S_1 and S_2 are iris diaphragms which can be opened and fully illuminated so as to provide an extended source for the initial setting up of the interferometer.

A rotatable aperture system A contains four pairs of circular apertures. The first is clear for both L_1 and L_2, the second places a defocusing spectacle lens in front of L_1, the third the same lens before L_2, and the fourth pair places defocusing lenses in front of both L_1 and L_2. In this way the frequency response of either or both lenses can be made zero.

The roof mirrors M and P are mounted on kinematic ways and can be accurately positioned along these by means of micrometer screws. The roof mirror P can also be moved in a direction parallel to the incoming beam and its position is adjusted to ensure roughly equal optical path lengths in the two arms of the interferometer. For incoherent illumination of the object the aperture I has a shape which masks off the non-interfering light from the sheared wave fronts, and so is merely the area common to two displaced circles. For partial coherence I could be further restricted by an aperture equivalent in size to the effective source.

The bed of the interferometer consists of a heavy Meehanite casting machined to take the kinematic ways of M and P and the mounting for D_1. These three components are very sensitive to external vibrations and to obtain reliable results a carefully designed anti-vibration support was found necessary. The casting is ribbed underneath and is supported on a thick cushion of Latex foam rubber placed on a marble slab, which is in turn mounted on four helical springs under compression. The efficiency of this system is such that under normal working conditions no movement of the fringes can be observed over a period of five minutes. There remains a slow movement with a period of about half an hour and this was assumed to be due to thermal changes.

The interferometer mirrors are fitted with 80 threads per inch screw adjustments and each is used in conjunction with a cam. The screw rests on this cam consisting of a ball bearing seated in a trihedral or conical hollow. The ball bearing can be rotated in this position by means of a rod to the end of which it is soldered, and owing to the residual asphericity of the ball its rotation imparts very small movements to the screw. The pressure on the ball bearing is adjusted by varying the tension in an opposing spring so as to ensure a regular fringe movement when set at its optimum value.

The variation of optical path difference between the two interfering beams has to be carried out with a precision of about $\lambda/100$ over about 1 λ and without introducing tilt. One way in which this was accomplished depended on the magnetostrictive effect of nickel. The mirror P was attached to one end of a carefully annealed nickel rod the other end of which was firmly fixed to the interferometer bed. By varying the current in a solenoid surrounding the rod the position of P could be adjusted as required. The advantage of this method is that by magnetically polarizing the rod and passing an alternating current through the coil this system will act as a magnetostrictive oscillator and, as we have seen, the amplitude of the flux output obtained with a given lens will be a measure of the modulus of its frequency response.

For some purposes, however, it is convenient to have direct control of the path differences and this was achieved by displacing the source off axis. Referring to figure 2 we see that if the beam were inclined at an angle β there would be a path difference between the two sheared beams. The distance from O through M to A is equal to that from O through P to E, so that the optical path difference is AB and is equal to the product of the shear EB and the angle β which is of course independent of the wavelength provided the lens is corrected for chromatic aberration. It is possible therefore to introduce optical path differences of the order required merely by displacing the slit S slightly off axis. It may be shown that in order to produce 1λ path difference between the interfering beams the slit must be displaced by a distance equal to the spatial frequency of the period under consideration.

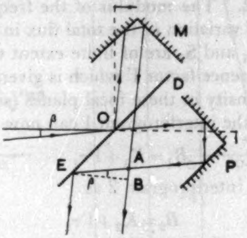


Figure 2. Provided there is a finite shear, path differences may be introduced by inclining the incoming wave as shown here.

§ 4. EXPERIMENTAL RESULTS

The apparatus was set up using the plane wave from L_1 so that the field was of uniform brightness for all values of the shear. The aberrant wave from L_2 was then positioned so as to take the same path through the interferometer, and this was done by adjusting for coincidence the exit pupils L_1 and L_2 imaged at I, and both the source images at K with G removed.

The position of M and P for zero shear could be accurately found since it is only here that the interferogram is independent of lens aberrations or defocusing. The shear was measured from this point with micrometer screws controlling M and P and was in error by about 0·1%.

The main reasons for the choice of a two-beam flicker photometer are firstly that by the use of a plane comparison wave it is possible to establish a direction of reference and so measure the argument of the frequency response, secondly that it is possible to compensate automatically for the finite source width at S_1 and S_2 by finding the relative frequency response of the two lenses, and lastly that a two-beam null method gives a result independent of source and detector fluctuations, and so is capable of a high degree of accuracy.

The total light flux in the interferogram of either L_1 or L_2 for a given δ was measured in each case by rotating A so as to place a defocusing lens in the wave front not being considered, which could then act as a standard photometric comparison beam. Thus if the unknown flux in interferogram 2 is B_2, that in the comparison arm containing the defocusing spectacle lens is ${}_{\infty}B_1$, and the setting of the polaroid P_1 resulting in no flicker at E_2 when P_2 is continuously rotated, is ϕ, then $B_2/{}_{\infty}B_1 = \tan^2\phi$.

For a particular shear the path difference was adjusted to give a maximum flux output of interferogram 1 corresponding to $\delta = 0$. The argument θ of L_2 can now be found as the lateral displacement of S_2 required to maximize the flux output of interferogram 2. The modulus of the frequency response τ_2 can be found as the amplitude of variation of the total flux in the interferogram 2.

If the slit images at S_1 and S_2 are of finite extent the visibility of the fringes will be reduced by a coherence factor Γ which is given by the Fourier transform of the distribution of intensity in these focal planes (see Hopkins 1951).

Thus the total flux in the interferogram 1 can now be written as

$$B_1 = K_1 + \Gamma\tau_1$$

and the maximum flux in interferogram 2 as

$$B_2 = K_2 + \Gamma\tau_2$$

where we have assumed that the source functions at S_1 and S_2 are identical, each giving the same phase coherence factor Γ.

If now we eliminate Γ from these two equations we are left with the ratio of the frequency response of the lens under test to that of the collimating lens in the form

$$\frac{\tau_2}{\tau_1} = \frac{B_2 - K_2}{B_1 - K_1} = \frac{B_2 - {}_{\infty}B_2}{B_1 - {}_{\infty}B_1}$$

which can be easily expressed in terms of the ratios $B_2/{}_{\infty}B_1$, $B_1/{}_{\infty}B_2$ and ${}_{\infty}B_1/{}_{\infty}B_2$, all of which can be found from angular settings of the polaroid P_1. Thus apart from the constant ${}_{\infty}B_1/{}_{\infty}B_2$ the relative frequency response can be determined at each shear by two settings of P_1.

In order to check the efficiency of the apparatus the frequency response of a defocused telescope objective was measured, and the results compared with those obtained theoretically (Hopkins 1955 b). The good quality comparison lens was an f/22 collimator and the lens under test was an f/22 good quality telescope doublet working on axis. In figure 3, $M(s)$ is the relative frequency response, that is the ratio of the response of the lens under test to that of the good quality lens. s is the shear of the wave fronts and has a maximum value of 2, and the wave aberration associated with the defocusing is $W_{20} = n\lambda/\pi$. Drawings of the sheared interferograms are shown at some values of s.

Figure 4 (*a*) shows the object frequencies just on the limit of resolving power for various values of the defocusing, and demonstrates how rapidly the resolved frequencies decrease for a small increase of the aberrations, and figure 4 (*b*) is a similar curve for the first minimum of the relative response. In all these graphs the full line curves are the theoretical ones and the points are experimental.

Since the total light flux available in most of these measurements was below one microlumen, giving a calculated field luminance of 350 candelas per square metre, it was difficult to obtain single results reproducible to better than 10% by visual means. The average of a number of independent settings of P_1, however, gave results which agreed to within about 5% of the theoretical curve. A convenient practical limit of the shear in these experiments was $s = 0{\cdot}2$. For larger values the source width had to be reduced below 0·004 in. for coherence, and

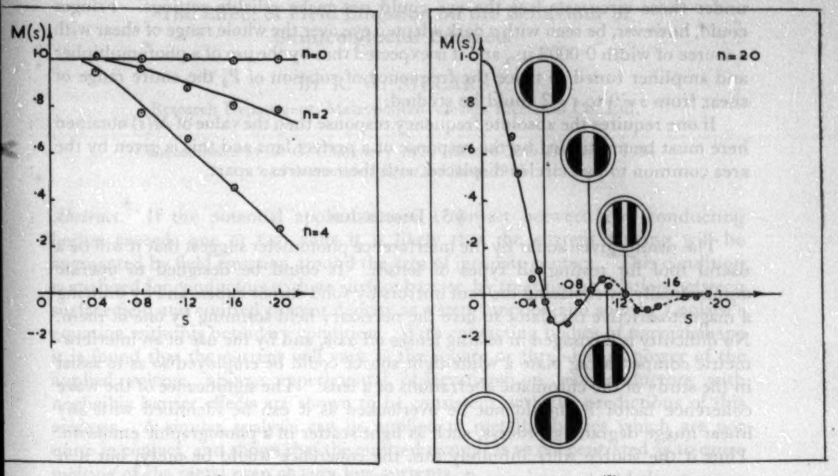


Figure 3. A plot of the relative frequency response $M(s)$ against the defocusing term s showing close agreement between theory and experiment. (*a*) $n=0, 2, 4$, (*b*) $n=20$.

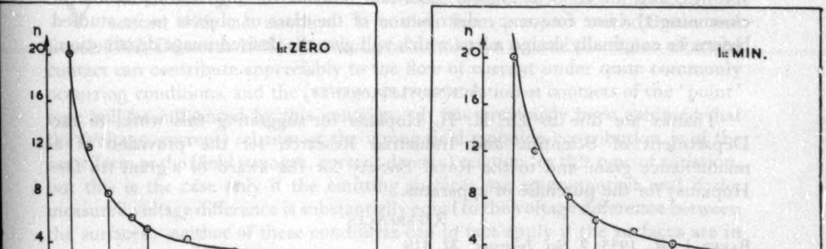


Figure 4. A plot of line frequency against the defocusing term for (*a*) the first zero and (*b*) the first minimum of the frequency response.

under these circumstances the eye could not make reliable settings. Fringes could, however, be seen with a dark adapted eye over the whole range of shear with a source of width 0·0002 in., and it is expected that by the use of a photomultiplier and amplifier tuned to twice the frequency of rotation of P_2 the entire range of shear from $s=0$ to $s=2$ could be studied.

If one requires the absolute frequency response then the value of $M(s)$ obtained here must be multiplied by the response of a perfect lens and this is given by the area common to two circles displaced with their centres s apart.

§ 5. Discussion

The results given so far by the interference photometer suggest that it will be a useful tool for testing all types of lenses. It could be designed to operate automatically by replacing the roof mirrors by solid corner cubes, and by utilizing a magnetostrictive oscillator to give the necessary field scanning by one of them. No difficulty is envisaged in testing lenses off axis, and by the use of an interferometric compensating plate a white light source could be employed so as to assist in the study of the chromatic aberrations of a lens. The significance of the wave coherence factor Γ should not be overlooked as it can be identified with any linear image degrading process, such as light scatter in a photographic emulsion. Thus if the source were infinitely thin the coherence would be unity, but if it were extended by transmission through a typical scattering emulsion the coherence would be reduced according to the transform of the scattering function, and the measured frequency response would then include the effect of a typical lens emulsion combination. In conclusion it should be emphasized that even if we can measure and control the frequency-response curve of a lens for all focal settings, and for all field angles, there still remains the important problem of examining the line frequency distribution of the class of objects to be studied before we can finally design a lens which will give the desired image distribution.

Acknowledgments

Thanks are due to Dr. H. H. Hopkins for suggesting this work, to the Department of Scientific and Industrial Research for the provision of a maintenance grant and to the Royal Society for the award of a grant (to Dr. Hopkins) for the purchase of apparatus.

References

Baker, L. R., 1955, *J. Sci. Instrum.*, **32**, 418.

Baker, L. R., and Hopkins, H. H., 1955, *U.K. Pat. Appl.* No. 9837/55.

Blanc-Lapierre, A., and Dumontet, P., 1955, *Rev. Opt.* (*théor. instrum.*), **34**, 1.

Duffieux, P. M., 1946, *L'intégrale de Fourier et ses applications à l'optique* (Besançon : privately printed).

Hardy, A. C., 1935, *J. Opt. Soc. Amer.*, **25**, 305.

Hopkins, H. H., 1951, *Proc. Roy. Soc.* A, **208**, 263; 1953, *Ibid.*, **217**, 408; 1955 a, Manchester Symposium on Astronomical Optics; 1955 b, *Proc. Roy. Soc.* A, **231**, 91; 1955 c, *Optica Acta*, **1**, 23.

Lindberg, P., 1954, *Optica Acta*, **1**, 80.

Rayleigh, Lord, 1896, *Phil. Mag.*, **42**, 167.

Selwyn, E. W. H., 1948 *Photogr. J.* B, **88**, 6.

Steel, W. H., 1952, *Rev. Opt.* (*théor. instrum.*), **31**, 334; 1953, *Ibid.*, **32**, 4, 143, 269.

Zernike, F., 1938, *Physica*, **5**, 785.

The Effect of Field Emission on the Behaviour of Semiconductor Contacts

By R. W. SILLARS
Research Department, Metropolitan-Vickers Electrical Co. Ltd.

Communicated by B. G. Churcher; MS. received 30th December 1954
and in amended form 2nd June 1955

Abstract. If the potential applied across a contact between two conducting bodies exceeds one or two volts it is likely that the current flowing will be augmented by field emission around the area of intimate contact. This condition is analysed for conductors with no surface barrier, by treating the relation between surface field and emitted current density as a step-function and solving Laplace's equation with this boundary condition. With contacting bodies of normal shape it is found that the current will vary as the square or three-halves power of the applied voltage. Some experimental observations on semiconductors with negligible barrier effects are shown to be consistent with the predictions of this analysis. A similar analysis can be applied to metallic bodies which are not quite in contact, and shows that the emitted current is so concentrated as to cause melting of the metal even at very low currents.

---

§ 1. INTRODUCTION

VARIOUS authors have discussed the passage of electrons, by tunnelling and by strong field emission, between conducting bodies imperfectly in contact with one another. It is clear from previous work (Braun and Busch 1942, Dilworth 1948) that strong field emission around the region of actual contact can contribute appreciably to the flow of current under quite commonly occurring conditions, and the (voltage, current) relation at contacts of the 'point' type will be influenced by this emission. It has previously been assumed that the (voltage, current) relation of the strong-field emission contribution is of the same form as the (field strength, current density) relation for this type of emission, but this is the case only if the emitting surfaces are equipotentials and if the measured voltage difference is substantially equal to the voltage difference between the surfaces; neither of these conditions can in fact apply if the surfaces are in contact at some point.

Tunnelling and field emission appeared at one time to be the only possible explanation of non-linear conduction at contacts between similar, clean semiconductors, such as is observed with silicon carbide; it now seems highly probable that a space-charge barrier exists at both surfaces due to the presence of surface traps for holes or electrons, and that many of the features of such contacts are associated with the passage of carriers over these barriers rather than with any form of tunnel effect. However, it is difficult to explain, on this picture alone, the very wide range of currents and voltages over which non-linear conduction persists, moreover it seems likely that at the higher values the barrier will scarcely be effective, so that a study of phenomena at contacts without such barriers is still of interest.

Finally, there are experimental observations of non-linear conduction, at contacts to semiconductors with a high density of impurity centres, where the voltage is such that it is scarcely conceivable that the space-charge barrier could produce any measurable effect; these seem incapable of explanation except by field emission around the contact region.

For these reasons a more careful consideration of this aspect of contacts is required, and the simplest case to consider is that of a conductor which is electrically homogeneous right up to the surface. Throughout the remainder of this paper the word semiconductor is used merely in the sense of a conductor of resistivity considerably higher than that of a metal, without implying any local change of conductivity near the surface.

In the following analysis, the possible limitation of emission by space charge (such as was postulated by v. Geel 1931) is ignored on the ground that it is unlikely to arise in most practical circumstances (Dyke and Trolan 1953).

§ 2. PRELIMINARY DISCUSSION

The relation between emitted current density I and field strength F is of the type

$$I = A \exp(-b/F) \qquad \ldots\ldots(1)$$

where A is either constant or varies relatively slowly with F (according to the assumptions made about the semiconductor and the approximations used), and b will be taken as constant for the present purpose. Commonly A has a value in the region of 10^6–10^9 A cm^{-2} under conditions of emission and b has a value of about 10^8–10^9 v cm^{-1}.† Stratton (1955) has given a detailed discussion of these quantities. Equation (1) is plotted in figure 1 with $A = 10^9$ A cm^{-2} and $b = 5 \times 10^8$ v cm^{-1} for two ranges of current density, to demonstrate that the field strength changes very slowly for a large change of current.

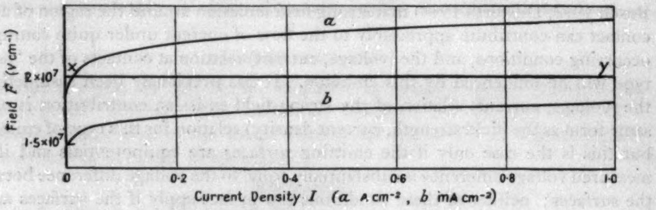


Figure 1. Typical field emission curve $I = 10^9 \exp(-5 \times 10^8/F)$.

If an ordinary resistance R is connected in series with a gap g, across which current is flowing according to equation (1), and if the total e.m.f. applied to the combination is v, the current flow will be $J = (v - gF)/R$, $(gF < v)$, where F is the quantity illustrated in figure 1 and varies very slowly with J or v. Evidently the variation of J with v is nothing like the variation of I with F in equation (1).

Consider now an idealized type of contact geometry illustrated in figure 2, which represents a cross section of the contact region, supposed to be symmetrical about the axis PQ. The potential of the semiconductor is the same as that of the perfect conductor, i.e. V_0 over the region of contact, radius *d*, and is supposed to be zero at infinite distance from it. This arrangement is substantially the same, for purposes of calculation, as that of figure 3(*a*) or 3(*b*), or half that of figure 3(*c*). The results will thus be applicable to a contact between two similar semiconductors, and likewise to each contact in a chain of semiconducting granules touching one another.

---

† There is now some evidence that, for conductors exposed to the atmosphere, the value of b may be lower than the value deduced from the ' clean ' work function by as much as one or two orders of magnitude. This point is considered in § 7.

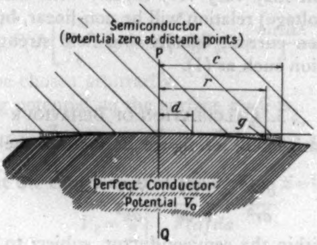

Semiconductor
(Potential zero at distant points)
P
c
r
d
g
Perfect Conductor
Potential V_0
Q

Figure 2. Idealized contact between semiconductor and metal.

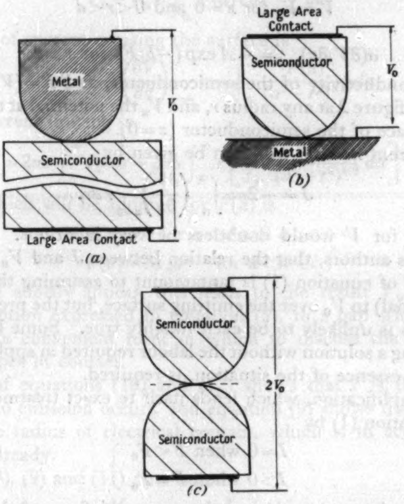

Metal
V_0
Semiconductor
Large Area Contact
(a)
Large Area Contact
Semiconductor
V_0
Metal
(b)
Semiconductor
$2V_0$
Semiconductor
(c)

Figure 3. Various types of contact which can be treated as in figure 2.

Suppose that, at a certain value of V_0, field emission is taking place over an area roughly limited by the radius *c* (figure 2). At points near the centre of this area the potential drop across the gap will be small compared with V_0, while the potential drop along the lines of current flow within the semiconductor leading to these points will be nearly equal to V_0. If we concentrate attention on this same region near the centre, while increasing V_0, most of the increase will be taken up within the semiconductor, and not across the gap, for the reasons detailed in the last paragraph but one. The contribution of current from this central area is therefore very nearly proportional to V_0. Now turning attention to the outer periphery of the emitting area, an increase of the total voltage V_0 will increase the width of the gap across which the field is great enough to cause appreciable emission, and will consequently increase the emitting area. This contribution of current may vary very rapidly with change of V_0. Evidently the overall (current, voltage) relation will be non-linear, but it will not be similar to the relation between current density and field strength, i.e. it will not be analogous to an equation such as (1).

§ 3. CALCULATION OF BEHAVIOUR

The idealized arrangement of figure 2 requires the solution of Laplace's equation

$$\frac{\partial^2 V}{\partial r^2}+\frac{1}{r}\frac{\partial V}{\partial r}+\frac{\partial^2 V}{\partial z^2}=0 \qquad \ldots\ldots(2)$$

for V the potential within the semiconductor, subject to the conditions

$$V=0 \text{ at } r=\infty \text{ and at } z=\infty$$

$$V=V_0 \text{ for } z=0 \text{ and } 0<r<d$$

and

$$\sigma(\partial V/\partial z)_{z=0}=-A\exp(-b/F) \text{ for } r>d \qquad \ldots\ldots(3)$$

where σ is the conductivity of the semiconductor, and $F=(V_s-V_0)/g$, g being the gap shown in figure 2 at any radius r, and V_s the potential at the corresponding point on the surface of the semiconductor ($z=0$).

The total current flowing will then be given by

$$J=-2\pi\sigma\int_0^\infty\left(\frac{\partial V}{\partial z}\right)_{z=0} r\,dr.$$

The solution for V would doubtless be very laborious. The assumption made by previous authors, that the relation between J and V_0 is similar to that between I and F of equation (1) is tantamount to assuming that V is equal (or at least proportional) to V_0 over the emitting surface, but the preceding discussion suggests that this is unlikely to be even roughly true. Some better simplifying assumption, giving a solution without the labour required in applying equation (3), yet retaining the essence of the situation, is required.

One such simplification, which lends itself to exact treatment subsequently, is to replace equation (1) by

$$\left.\begin{array}{l} I=0 \text{ when } F<F_0 \\ I>0 \text{ when } F=F_0 \end{array}\right\} \qquad \ldots\ldots(4)$$

This is equivalent to replacing the curve of figure 1 by the rectilinear characteristic OXY where the value of F at X is F_0. The field over the emitting area is then constant and Laplace's equation within the gap is approximately satisfied by

$$V_0-V_s=gF_0 \qquad \ldots\ldots(5)$$

which can be used instead of equation (3) as the boundary condition for the potential within the semiconductor. Evidently with this approximation emission will cease sharply at a certain radius. The potential will not change abruptly, however, so that equation (5) will not be seriously in error even at this radius provided the surfaces are still nearly parallel.

The general solution of equation (2) (see Carslaw and Jaeger 1947) is

$$V = \text{const.} \int_0^\infty \exp(-\lambda |z|) J_0(\lambda r) f(\lambda)\, d\lambda$$

where $f(\lambda)$ is an arbitrary function. A convenient form of solution for the present purpose is

$$V = \int_0^\infty F(a) \left[\int_0^\infty \exp(-\lambda |z|) J_0(\lambda r) \frac{\sin a\lambda}{\lambda}\, d\lambda \right] da \qquad \ldots\ldots(6)$$

where $F(a)$ can also be chosen arbitrarily.

The corresponding potential on the surface $z=0$ is

$$V_s = \int_0^r F(a) \sin^{-1}\frac{a}{r}\, da + \frac{\pi}{2}\int_r^\infty F(a)\, da \qquad \ldots\ldots(7)$$

so that the potential at the centre $r=0$ of the surface $z=0$ is

$$V_0 = \frac{\pi}{2}\int_0^\infty F(a)\, da \qquad \ldots\ldots(8)$$

giving

$$V_0 - V_s = \int_0^r F(a) \cos^{-1}\frac{a}{r}\, da. \qquad \ldots\ldots(9)$$

The density of current crossing the surface is

$$I_s = -\sigma\left(\frac{\partial V}{\partial z}\right)_{z=0} = \sigma\int_r^\infty \frac{F(a)}{(a^2-r^2)^{1/2}}\, da \qquad \ldots\ldots(10)$$

and the total current flowing is

$$J = 2\pi\sigma \int_0^\infty I_s r\, dr = 2\pi\sigma \int_0^\infty r \left[\int_r^\infty \frac{F(a)}{(a^2-r^2)^{1/2}}\, da \right] dr. \qquad \ldots\ldots(11)$$

The form which will be adopted for $F(a)$ is

$$\left.\begin{aligned} F(a) &= 0 \text{ for } a<d \text{ and } a>c \\ F(a) &= M_n a^{n-1} \text{ for } d<a<c \end{aligned}\right\} \qquad \ldots\ldots(12)$$

because, as will shortly appear, this leads to the result $V_0 - V_s \propto r^n$ and so, with the assumption expressed in equation (5), to a gap width g proportional to r^n which is a convenient form in which to discuss the effect of different geometrical shapes of contact.

Inspection of equations (10) and (12) shows that c is the limiting radius beyond which no emission occurs, and equation (9) shows that $V_s = V_0$ for $r<d$ so that d is the radius of electrical contact, which is in accordance with the notation used already.

Equations (8), (9) and (11) now become

$$V_0 = \frac{M_n \pi}{2n}(c^n - d^n) \qquad \ldots\ldots(8a)$$

$$V_0 - V_s = M_n \int_d^r a^{n-1} \cos^{-1}\frac{a}{r}\, da \quad (d<r<c) \qquad \ldots\ldots(9a)$$

$$V_0 - V_s = M_n \int_d^c a^{n-1} \cos^{-1}\frac{a}{r}\, da \quad r \geqslant c \qquad \ldots\ldots(9b)$$

$$J = 2\pi\sigma M_n \int_0^c r \left[\int_r^c \frac{a^{n-1}}{(a^2-r^2)^{1/2}}\, da \right] dr. \qquad \ldots\ldots(11a)$$

3.1. *Vanishingly Small Area of Contact*

The general character of the results is most easily seen from the simple case where the area of contact is much less than the area over which emission occurs, ($d^2 \ll c^2$) so that the lower limit of integration can be replaced by zero. Equations (8*a*) and (9*a*) then become respectively

$$V_0 = \frac{M_n \pi}{2n} c^n$$

and

$$V_0 - V_s = \frac{M_n}{2n} \mathrm{B}\left(\frac{1}{2}, \frac{n+1}{2}\right) r^n = \frac{M_n \sqrt{\pi}}{2n} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma[\frac{1}{2}(n+2)]} r^n$$

where B( ) and Γ( ) represent the beta and gamma (factorial) functions.

The last equation evidently enables equation (5) to be satisfied if the gap g varies with r according to

$$g = K_n r^n \qquad \ldots\ldots(13)$$

with

$$\frac{M_n}{2n} \mathrm{B}\left(\frac{1}{2}, \frac{n+1}{2}\right) = K_n F_0.$$

Equations (8*a*), (9*a*) and (11*a*)† then become, after eliminating M_n

$$V_0 = \frac{\pi K_n F_0}{\mathrm{B}[\frac{1}{2}, \frac{1}{2}(n+1)]} c^n \qquad \ldots\ldots(14)$$

$$V_0 - V_s = K_n F_0 r^n \quad (r < c) \qquad \ldots\ldots(15)$$

(there is no need to evaluate the integral of equation (9*b*) for the present purpose), and

$$J = \frac{n}{n+1} 4\sigma V_0 c \qquad \ldots\ldots(16)$$

$$= \frac{n}{n+1} 4\sigma \left(\frac{\mathrm{B}[\frac{1}{2}, \frac{1}{2}(n+1)]}{\pi K_n F_0}\right)^{1/n} V_0^{1+1/n}. \qquad \ldots\ldots(17)$$

The equations (13) to (17) show that if the gap distance g is proportional to r^n, the radius of emission is proportional to $V_0^{1/n}$ or to $(J/\sigma)^{1/(n+1)}$ or to $J/\sigma V_0$ while the current is proportional to $(\text{voltage})^{1+1/n}$. To appreciate these equations better, the table gives the numerical relationships for three types of contact tip: spherical or parabolic ($n=2$), conical ($n=1$), and a concave-sided point ($n=\frac{1}{2}$). Evidently the (voltage, current) relation for any commonly expected type of contact geometry, while not linear, will be much less non-linear than the field emission relationship of equation (1). It is extremely unlikely that the (voltage, current) relation will be greatly affected by the precise form of the field emission relationship, or that the use of the correct field emission equation in place of the step-function approximation (4) would make this relation more non-linear.

It should be remembered that the equations are derived neglecting the inclination between the surfaces so that for the spherical tip they are only accurate within a radius much less than the radius of the sphere, for the conical tip only when the cone is of large angle, while for a concave-sided point they apply only when the radius of emitting surface is large enough for the majority of the surfaces within it to be nearly parallel.

It is worth noting that, with a spherical tip radius of curvature R, the current passed is the same as would be passed through a contact of radius

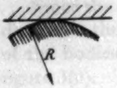
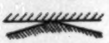
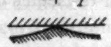
$$s = \tfrac{2}{3}(R V_0 / F_0)^{1/2} = \tfrac{2}{3} c$$

---

† Equation (11*a*) may be integrated by first inverting the double integral (see Whittaker and Watson 1920) or by expanding the inner integral in ascending powers of r/a.

i.e. two thirds of the actual emitting radius. In the case of a conical contact the corresponding value is half the emitting radius. Also, if $F_0 = 2 \times 10^7$ v cm^{-1}, and R is 10^{-3} cm the emitting radius will be 2×10^{-5} cm (of the same order as a 'catswhisker' contact radius) at 8 volts.

Table of Relations for Vanishingly Small Contact Area
(from equations (14), (16) and (17))

Type of contact	n	Radius of emitting area	(Volt, current) relation
Gap = $\frac{r^2}{2R}$ 	2	$c=\left(\frac{RV_0}{F_0}\right)^{1/2}=\frac{1}{2}\left(\frac{3RJ}{\sigma F_0}\right)^{1/3}$	$J=\frac{8\sigma}{3}\left(\frac{R}{F_0}\right)V_0^{3/2}$
Gap = $\frac{r}{S}$ 	1	$c=\frac{2SV_0}{\pi F_0}=\left(\frac{SJ}{\pi\sigma F_0}\right)^{1/2}$	$J=\frac{4\sigma S}{\pi F_0}V_0^2$
Gap = $\frac{\sqrt{r}}{T}$ 	$\frac{1}{2}$	$c=0{\cdot}582\left(\frac{TV_0}{F_0}\right)^2$ $=0{\cdot}689\left(\frac{TJ}{\sigma F_0}\right)^{2/3}$	$J=0{\cdot}776\,\frac{\sigma T^2}{F_0^2}V_0^3$

3.2. *Finite Contact Region*

To avoid complexity only the case of a spherical tip ($n=2$) will be considered. Equations (8*a*), (9*a*) and (11*a*), putting n equal to 2 and retaining d, become

$$V_0 = \frac{M_2\pi}{4}(c^2-d^2)$$

$$V_0 - V_s = \frac{M_2 r^2}{4}\left[\left(1-2\frac{d^2}{r^2}\right)\cos^{-1}\frac{d}{r}+\frac{d}{r}\left(1-\frac{d^2}{r^2}\right)^{1/2}\right], \quad d<r<c$$

$$J = \frac{2\pi\sigma M_2}{3}(c^2-d^2).$$

To satisfy equation (5) we may now put

$$M_2 = 4F_0/\pi R$$

and

$$g = \frac{r^2}{\pi R}\left[\left(1-2\frac{d^2}{r^2}\right)\cos^{-1}\frac{d}{r}+\frac{d}{r}\left(1-\frac{d^2}{r^2}\right)^{1/2}\right]. \qquad \ldots\ldots(18)$$

This expression for g is shown graphically in figure 4 to demonstrate that it is a plausible imitation of a sphere of radius R pressed against a plane to form a contact circle of radius d. Combination of these with the preceding equations gives $V_0 - V_s = gF_0$, which is equation (5), as intended, and

$$c^2 = V_0 R/F_0 + d^2 \qquad \ldots\ldots(19)$$

$$J = \frac{8}{3}\frac{\sigma F_0 d^3}{R}\left[\left(1+\frac{V_0 R}{F_0 d^2}\right)^{3/2}-1\right]. \qquad \ldots\ldots(20)$$

For large values of the term in the bracket, equation (20) gives the same result as already obtained when $d=0$ (see table of relations), for small values it reduces to

$$J=4\sigma\, dV_0\left\{1+\frac{1}{4}\frac{V_0R}{F_0d^2}+\ldots\right\} \qquad \ldots\ldots(21)$$

i.e. the conductance J/V_0 increases linearly with applied voltage.

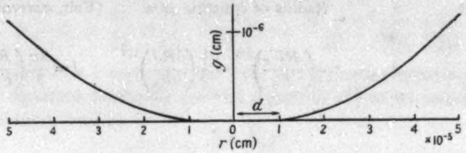


Figure 4. Curve of g from equation (1b) with $R=10^{-3}$ cm, $d=10^{-5}$ cm.

In the absence of field emission the current passing would of course be $4\sigma\, dV_0$ so that field emission doubles the current flowing when

$$3\,\frac{V_0R}{F_0d^2}=\left(1+\frac{V_0R}{F_0d^2}\right)^{3/2}-1$$

i.e. when $V_0R/F_0d^2\simeq 7$ or $V_0=14\times 10^7\, d^2/R$ volts if $F_0=2\times 10^7\ \mathrm{v\,cm^{-1}}$.

If $d=10^{-5}$ cm and $R=10^{-3}$ cm an applied potential of 14 volts will cause a current twice as great as would be calculated from the constriction resistance of the actual electrical contact.

§ 4. CONDUCTORS NOT QUITE IN CONTACT

The equations derived in § 3.1 can be applied with slight modifications to the arrangement of figure 5, consisting of two conductors 1 and 2, with spherical tips, separated by a gap $2g_0$ which is much smaller than the radius of curvature. The potential at remote points in conductor 1 will be taken as zero and that at remote points in conductor 2 will be called $2V_c$, so that the plane of symmetry in figure 5 will be the equipotential V_c. The potentials at the vertices of the two conductors will be called V_1 and V_2 so that

$$V_c=V_1+g_0F=V_2-g_0F.$$

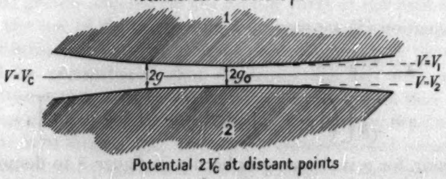

Potential zero at distant points
1
$V=V_1$
$V=V_c$
$2g$
$2g_0$
$V=V_2$
2
Potential $2V_c$ at distant points

Figure 5. Conductors separated by a small gap.

The equipotentials in the space between the conductors and within the emitting region are equally spaced planes parallel to the central equipotential V_c. Those corresponding to potentials less than V_1 or greater than V_2 intersect the conductors 1 and 2 respectively, while the equipotentials V_1 and V_2 are tangential to them. This is obviously consistent with equation (5). It is also consistent with the analysis applied in the earlier part of this paper, for the field normal to the surface of the conductor was, by hypothesis, F_0 and the field tangential to it dV_s/dr was, by equation (5), equal to $-F_0\,dg/dr$, so that the resultant field makes an angle $\tan^{-1}(dg/dr)$ with the normal to the surface, i.e. it is parallel to the axis of the figure and the equipotentials are therefore perpendicular to the axis. (There is, of course, an inconsistency of the second order in that we implied at one point in the argument that the field strength is F_0, and at another point that it is $F_0\{1+(dg/dr)^2\}^{1/2}$, but this inconsistency can be neglected so long as the inclination between the surfaces is small.)

Since the equipotentials V_1 and V_2 are (in the emitting region) planes perpendicular to the axis and tangential to the vertices of the conducting bodies, each of the bodies is in the same case as is shown for the semiconducting body of figure 3(*b*). The equations of § 3.1 can be applied to conductor 1 (or 2) with V_1 (or $2V_c-V_2$) substituted for V_0 and $g-g_0$ for g. If the conductors are spheres of radius R it follows that the (voltage, current) relation is

$$J=\frac{8\sigma}{3}\left(\frac{R}{F_0}\right)^{1/2}V_1^{3/2}=\frac{8\sigma}{3}\left(\frac{R}{F_0}\right)^{1/2}(V_c-g_0F_0)^{3/2}.$$

The radius of the emitting region will be $c=\frac{1}{2}(3RJ/\sigma F_0)^{1/3}$ as before.

It is interesting to insert values appropriate to a metal in the expression for c. Putting $\sigma=2\times10^5\,(\Omega\text{ cm})^{-1}$, $F_0=2\times10^7\text{ v cm}^{-1}$ and taking R as 10^5 cm gives c as 2×10^{-3} cm for a current of 1 ampere. Thus even with very flat plates the current confines itself to a very narrow channel. The electrons crossing the gap will bombard the anode with an energy density approximately given by $I_s g_0 F_0=Q_s$. Since Q_s has the same distribution as I_s it follows that the central temperature T_0 is related to the total heat flow Jg_0F_0 in the same way as the central voltage was related to the total current in equation (16). Putting $n=2$ and making appropriate adaptations in this equation gives $Jg_0F_0=8\lambda T_0 c/3$ where λ is the thermal conductivity; and substituting for c gives

$$T_0=\frac{g_0}{4\lambda}\left(\frac{9\sigma F_0^{\,4}J^2}{R}\right)^{1/3}=3\times10^9 g_0 J^{2/3}$$

with $\lambda=1\text{ w cm}^{-1}\text{ deg}^{-1}$ and the other quantities as before.

So with a gap of only 10^{-4} cm the temperature would exceed the melting point of a metal with a current of the order of a milliampere. If the electrodes had a marked curvature, say $R=1$ mm instead of $R=10^5$ cm, a current of a few microamperes would suffice to produce local melting of the anode.

§ 5. THE EXTENT OF ELECTRICAL CONTACT

The region of electrical contact between two bodies has been referred to throughout this paper without any precise definition. It obviously includes the finite region which supports mechanical pressure, and in which the distance between constituent ions or atoms across the surface (in the absence of any intermediate surface layer) is about the same as within the bodies. Surrounding this region will be one in which the surfaces of the two bodies are still within a few ångströms of each other, and since a gap of this order of size is transparent to electrons, independent of the applied potential, this region is included in the electrical contact area. Outside this again is the region which the preceding analysis treats on the assumption that it suddenly becomes transparent when the surface field reaches $F_0 \simeq 2 \times 10^7\,\mathrm{v\,cm}^{-1}$. The two latter regions overlap at points where the potential difference between the surfaces is about 1 volt or less, and the results of this paper are irrelevant wherever this is the case. Provided the total applied potential exceeds a few volts, this 'overlap' region, which can neither be regarded as one of electrical contact nor as one where field emission relations apply, will be but a small fraction of the whole, because the transparency of the gap is a very critical function of its width.

It follows that this analysis cannot be applied to good conductors in contact since it is impossible to apply potentials of a few volts between them, but there is no reason why it should not be applied to good conductors not quite in contact, as in the last section. With most semiconductors, on the other hand, contacts with several volts potential difference between the contacting bodies are quite common.

For the purpose of this paper, then, the term 'electrical contact' refers to the region in which the surfaces of the contacting bodies are within a few ångströms of each other.

§ 6. APPLICATION TO EXPERIMENTAL RESULTS

The (voltage, current) relations of contacts involving semiconductors are influenced by so many factors that it is rarely possible to isolate a single one in any experiment. The effect considered here will appear in its simplest form with a semiconductor in which surface barrier effects are not noticeable, at least when the potential difference across the contact exceeds a few volts. This is the case with several conducting oxides, and estimates of the density of ionizable impurity centres indicate that this is probably because the normally expected space-charge barrier is too thin and too patchy to be effective.

One of these, magnesium ortho-titanate, has been the subject of investigations with tungsten 'point' contacts (Stuckes 1953) to observe the effects of self-heating at the contact region. The (voltage, current) curves displayed a curvature at voltages too low to be explained by heating effects, and too high (10–50 volts) to be explained by any probable space-charge barrier. An example is given in figure 11 of Stuckes' paper, which corresponds with contact 24*b* of table 3 of the same paper, where the low-voltage resistance of the contact is given as $1{\cdot}6 \times 10^8$ ohms. This value and the mean of the two curves of figure 11 up to 50 volts can be closely fitted by equation (20) of the present paper, putting $R/F_0d^2 = 0{\cdot}22$ and $3R/8\sigma F_0 d^3 = 5{\cdot}3 \times 10^7$. The contact radius d is of the order of 10^{-5} cm (see below) so that with $F_0 = 2 \times 10^7\,\mathrm{v\,cm}^{-1}$ the radius R of the etched tungsten point would be about 5×10^{-4} cm which is not an unexpected value.

The thermally produced peaks in figure 11 of Stuckes' paper occur at about 180 volts, and at this value the ratio of emitting radius to electrical contact radius c/d from equation (19) of the present paper is 7·3, suggesting that the 'contact radius' deduced from this peak will be larger than that appropriate to low voltage measurements, by a factor greater than unity but probably less than 7. This accords with a very noticeable feature of Stuckes' table 3 (column 7), that the ratio of resistance at low voltage to resistance at peak voltage is always higher than the purely thermal theory predicts, generally by a factor lying between 2 and 10. The data for contact 24 *b* can be manipulated in various ways to give values for the electrical contact radius (*d* of the present paper); perhaps the most reliable is deduced from the peak voltage (which gives a value for the local resistivity) and the constriction resistance of $1{\cdot}6 \times 10^8$ ohms; this gives $d = 0{\cdot}8 \times 10^{-5}$ cm. The values of 'contact radius' deduced from the peak conditions are about 2 and 3 times this, which is not wildly inconsistent with $c/d = 7$.

At the other extreme of geometrical shape are some unpublished observations also by Stuckes, on the current flowing between two optically flattened discs, of metal and magnesium titanate respectively, when laid together. The current generally varies as the square or three-halves power of the voltage, presumably depending on the detailed shape of the surface. Some curves showing the three-halves power dependence are given in figure 6. The discs are about 15 mm in diameter and their radii of curvature are known, from observations of interference fringes, to be of the order of 5000 cm; assuming them to be roughly spherical the radius of contact under the conditions of the experiment could be estimated from the elastic modulus using Herz's law. This estimate showed that the term $(V_0 R/F_0 d^2)^{3/2}$ in equation (20) would be large, so that the simple relations given in the top row of the table of relations derived from equation (17) could be used, and the expected (current, voltage) relation is

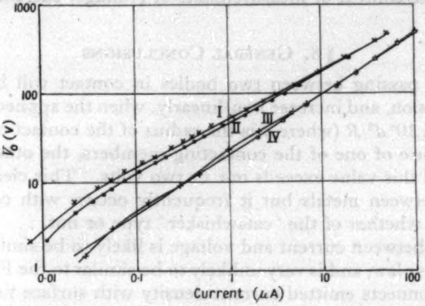


Figure 6. (Voltage, current) curves of metal and semiconductor discs placed together.

$$J = \frac{8\sigma}{3} \sqrt{\frac{R}{F_0}}\, V_0^{\,3/2}.$$

Curves I, II, III and IV of figure 6 can be roughly fitted by $J/V_0^{\,3/2} = 2{\cdot}8 \times 10^{-9}$, $3{\cdot}3 \times 10^{-9}$, $1{\cdot}1 \times 10^{-8}$ and $1{\cdot}2 \times 10^{-8}$ respectively. Curves I and II refer to a disc of bulk conductivity $4{\cdot}8 \times 10^{-8}\ (\Omega\ \mathrm{cm})^{-1}$, curves III and IV to a disc of conductivity $2{\cdot}8 \times 10^{-7}\ (\Omega\ \mathrm{cm})^{-1}$; with these figures, and assuming $F_0 = 2 \times 10^7\ \mathrm{v\ cm}^{-1}$ the values found for *R* are $1{\cdot}0 \times 10^3$ and $1{\cdot}4 \times 10^3$ cm for one disc, and $4{\cdot}4 \times 10^3$ and $5{\cdot}2 \times 10^3$ cm for the other, in tolerable agreement with the observed degree of flatness of the discs.

§ 7. VALUE OF CRITICAL FIELD F_0

It has been assumed that the best value for F_0 is about $2 \times 10^7\ \mathrm{v\ cm}^{-1}$, a value appropriate to a clean metal emitting in vacuum. Experiments on the initiation of breakdown across very small gaps, however, have led to the suggestion that emission may occur at much lower field strength under normal atmospheric exposure. Atalla (1953) found that this happens unless the electrodes have been 'cleaned' by severe arcing. Llewellyn Jones and de la Perrelle (1953) concluded that the emitted current involved is very small (10^{-14} A or less) and comes from a minute area of low work function such as a single adsorbed ion.

The analysis in this paper, using a suitable value of F_0, would apply to any conductor of fairly uniform work function, and qualitatively to a uniform distribution of emitting spots all having the same low work function. Minor variations would not invalidate it greatly because the local constriction resistance would limit the supply of current to the areas of slightly lower work function. The analysis would not be relevant to emission from individual highly localized regions, nor would the current from a region of a few ångströms diameter be appreciable in the context of measurements of (voltage, current) characteristics.

§ 8. General Conclusions

The current passing between two bodies in contact will be augmented by strong field emission, and increase non-linearly, when the applied voltage becomes comparable with $10^8 d^2/R$ (where d is the radius of the contact circle and R is the radius of curvature of one of the contacting members, the other being assumed planar) provided this value exceeds one or two volts. This clearly cannot occur with contacts between metals but it frequently occurs with contacts involving semiconductors, whether of the 'catswhisker' type or not.

The relation between current and voltage is likely to be similar to a square or three-halves power law, and is very unlikely to be similar to the Fowler–Nordheim relation which connects emitted current density with surface field strength.

A similar (voltage, current) relation will apply to conducting bodies, including metallic conductors, when separated by a very small gap. It is shown that when both bodies are metals the current is confined to a very small area, and the heat developed by electrons impinging on the anode is sufficiently concentrated to melt the metal even at currents considerably less than a milliampere.

These conclusions are reached by substituting a step-function relation for the Fowler–Nordheim or other similar relation, so that field emission from a body is treated as suddenly occurring at a threshold energy F_0; this field is then regarded as independent of emitted current density. The best value to choose for F_0 will, of course, depend on the form and parameters of the field emission equation, which may vary somewhat according to the assumptions used in deriving it (see Stratton 1955) and will differ with different materials. In the case of a metal–semiconductor contact, for instance, some degree of rectification is to be expected; but the little known effects of adsorbed layers on surface conditions are factors which preclude any general prediction of the preferred direction of flow under normal experimental conditions.

Acknowledgments

The author wishes to thank Miss A. D. Stuckes for many observations, Dr. R. Stratton for much detailed discussion of the problem, Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

ATALLA, M. M., 1953, *Bell Syst. Tech. J.*, **32**, 1493.

BRAUN, A., and BUSCH, G., 1942, *Helv. Phys. Acta*, **15**, 571.

CARSLAW, H. S., and JAEGER, J. C., 1947, *Conduction of Heat in Solids* (Oxford : Clarendon Press), p. 187.

DILWORTH, C. C., 1948, *Proc. Phys. Soc.*, **60**, 315.

DYKE, W. P., and TROLAN, J. K., 1953, *Phys. Rev.*, **89**, 799.

v. GEEL, W. CH., 1931, *Z. Phys.*, **69**, 765.

LLEWELLYN JONES, F., and DE LA PERRELLE, E. T., 1953, *Proc. Roy. Soc.* A, **216**, 267.

STRATTON, R., 1955, *Proc. Phys. Soc.* B, **68**, 746.

STUCKES, A. D., 1953, *Proc. Phys. Soc.* B, **66**, 570.

WHITTAKER, E., and WATSON, G. N., 1920, *Modern Analysis* (Cambridge : University Press), p. 77.

An Easy Derivation of the Hole Lifetime in an n-type Semiconductor with Acceptor Traps

BY F. W. G. ROSE AND D. J. SANDIFORD
Research Laboratory, British-Thomson-Houston Co. Ltd., Rugby

MS. received 5th April 1955 *and in amended form 26th May* 1955

Abstract. The recombination and generation processes for electrons and holes in an n-type semiconductor with acceptor traps are described when the concentrations of electrons and holes are much smaller than $10^{19}\,\mathrm{cm}^{-3}$ at room temperature. With this assumption Boltzmann statistics and the mass-action laws apply. The relations between the constants of recombination, generation and the mass-action laws are described. These relations are used for the calculation of lifetime in the case of steady state injection. It is shown that instantaneous lifetime and steady state lifetime are different. Agreement with an equation by Shockley and Read is found.

---

§ 1. INTRODUCTION

As shown by Ransom and Rose (1954) the equation (A 7) of Shockley and Read (1952) for the steady state lifetime of holes in an n-type semiconductor can be written

$$\tau_{\mathrm{p}}=\frac{(\alpha_{\mathrm{n}}N)^{-1}(p_0+p_1)+(\alpha N)^{-1}[n_0+n_1+N(1+n_0/n_1)^{-1}]}{n_0+p_0+N(1+n_0/n_1)^{-1}(1+n_1/n_0)^{-1}} \qquad \ldots\ldots(1)$$

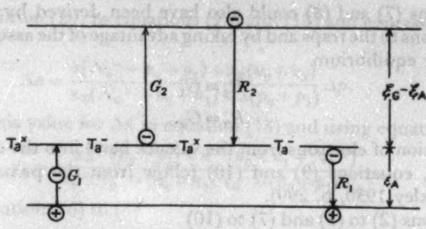
if acceptor traps are assumed. In this equation n_0 and p_0 are the equilibrium concentrations of free electrons and holes respectively, α_{n} and α are recombination constants for the trapping of free electrons or free holes by neutral and charged traps respectively. N is the concentration of traps, n_1 and p_1 are constants of the mass-action laws for the trapping processes of free electrons and free holes respectively and are later defined by equations (7) and (8).

The purpose of this publication is to show that equation (1) can be derived in an easy way when the concentrations are so small that Boltzmann statistics apply, i.e. when n_0 is much smaller than the effective density N_{c} of the energy levels in the conduction band and when p_0 is much smaller than the effective density N_{v} of energy levels in the valence band (Shockley 1950, pp. 240–242).

§ 2. RECOMBINATION AND GENERATION PROCESSES

In the figure the possible reactions between free electrons, free holes and acceptor traps are shown.

In this figure R_1 is the rate with which electrons drop from negatively charged traps into holes (trapping rate of free holes). R_2 is the rate of recombination of electrons in the conduction band with neutral acceptor traps (trapping rate of free electrons). G_1 is the rate with which electrons are lifted up from the valence band into neutral acceptor traps (generation rate of free holes). G_2 is the rate of



G_1 is the rate of generation of holes by neutral acceptor traps $T_a^\times$, G_2 the rate of generation of free electrons by charged acceptor traps T_a^-, R_1 the rate of trapping of holes by charged acceptor traps T_a^- and R_2 the rate of trapping of free electrons by neutral acceptor traps $T_a^\times$.

thermal excitation of electrons from negatively charged traps into the conduction band (generation rate of free electrons).

The following equations apply:

$$R_1 = \alpha p N^- \qquad \ldots\ldots(2)$$

$$G_1 = b_1 N^\times \qquad \ldots\ldots(3)$$

$$R_2 = \alpha_n n N^\times \qquad \ldots\ldots(4)$$

$$G_2 = b_2 N^- \qquad \ldots\ldots(5)$$

where n, p, N^-, $N^\times$ are the concentrations of free electrons, free holes, negatively charged traps and neutral traps respectively. The corresponding concentrations for the case of equilibrium will be marked with a subscript 0. The recombination constants α and α_n are defined by equations (2) and (4). Equations (3) and (5) define the temperature dependent coefficients b_1 and b_2. The total concentration of acceptor traps is

$$N = N^- + N^\times. \qquad \ldots\ldots(6)$$

The trapping of holes by ordinary donors is neglected in the following. Such donors are situated very near to the conduction band and therefore they have practically all lost their electrons to this band by thermal excitation, except at very low temperatures. Thus they cannot trap holes. The effect of a direct recombination of free electrons and free holes has also been neglected on account of the small capture cross section which must be assumed for electrons and holes.

If the mass-action laws are applied to the processes

$$T_a^\times \rightleftharpoons T_a^- + \text{hole}$$
$$T_a^- \rightleftharpoons T_a^\times + \text{electron}$$

of the figure, then the following equations apply for the case of equilibrium

$$\frac{p_0 N_0^-}{N_0^\times} = p_1 \equiv N_v \exp\left(\frac{-\mathscr{E}_A}{\boldsymbol{k}T}\right) \qquad \ldots\ldots(7)$$

$$\frac{n_0 N_0^\times}{N_0^-} = n_1 \equiv N_c \exp\left[-\frac{(\mathscr{E}_G - \mathscr{E}_A)}{\boldsymbol{k}T}\right] \qquad \ldots\ldots(8)$$

where $\mathscr{E}_A$ is the energy difference between the trap level and the valence band and where $\mathscr{E}_G$ is the energy gap. T is the temperature in degrees Kelvin and $\boldsymbol{k}$ is Boltzmann's constant. The use of the mass-action laws in non-degenerate electronic processes is well established (e.g. Schottky and Spenke 1939, Schottky

1952). Equations (7) and (8) could also have been derived by applying Fermi weighting functions to the traps and by taking advantage of the assumed Boltzmann conditions. For equilibrium

$$R_1 = G_1 \qquad \text{......(9)}$$

$$R_2 = G_2. \qquad \text{......(10)}$$

If a direct transition of electrons from the valence band into the conduction band is not neglected, equations (9) and (10) follow from the principle of detailed balancing (Shockley 1950, p. 299).

From equations (2) to (5) and (7) to (10)

$$p_1 = b_1/\alpha \qquad \text{......(11)}$$

$$n_1 = b_2/\alpha_{\mathrm{n}}. \qquad \text{......(12)}$$

§ 3. LIFETIME

The lifetime of a surplus concentration of holes is (Shockley 1950, p. 298), by definition

$$\tau_{\mathrm{p}} = \frac{\Delta p}{\Delta(R_1 - G_1)} \qquad \text{......(13)}$$

where Δ indicates a change from the thermal equilibrium value.

From equations (2) and (3), neglecting second order terms,

$$\begin{aligned} \Delta(R_1 - G_1) &= \alpha p_0 \Delta N^- + \alpha N_0^- \Delta p - b_1 \Delta N^\times \\ &= \alpha p_0 \Delta N^- + \alpha N_0^- \Delta p - \alpha p_1 \Delta N^\times \end{aligned} \qquad \text{......(14)}$$

substituting for b_1 from equation (11).

From equation (6),

$$\Delta N^- = -\Delta N^\times \qquad \text{......(15)}$$

since N is constant.

From equations (14) and (15)

$$\Delta(R_1 - G_1) = \alpha N_0^- \Delta p - \alpha(p_0 + p_1)\Delta N^\times. \qquad \text{......(16)}$$

In an extrinsic n-type semiconductor where a very high concentration of electrons is available the electrons will easily move in such a way that neutrality is reached, therefore

$$\Delta p = \Delta n + \Delta N^- = \Delta n - \Delta N^\times. \qquad \text{......(17)}$$

Substitution from equation (17) in equation (16) gives

$$\Delta(R_1 - G_1) = \alpha(N_0^- + p_0 + p_1)\Delta p - \alpha(p_0 + p_1)\Delta n. \qquad \text{......(18)}$$

Similarly

$$\Delta(R_2 - G_2) = \alpha_{\mathrm{n}}(N_0^\times + n_0 + n_1)\Delta n - \alpha_{\mathrm{n}}(n_0 + n_1)\Delta p. \qquad \text{......(19)}$$

3.1. *Steady State Lifetime*

In cases when holes are injected and a steady excess hole concentration is maintained (this applies for instance to lifetime measurements by the Morton–Haynes method, Valdes 1952) the concentrations of charged and uncharged acceptor traps respectively must reach a steady value. In this case

$$\Delta(R_1 - G_1) = \Delta(R_2 - G_2). \qquad \text{......(20)}$$

Applying condition (20) to equations (18) and (19) and expressing in terms of Δn and Δp,

$$\Delta n = \frac{\alpha(N_0^- + p_0 + p_1) + \alpha_n(n_0 + n_1)}{\alpha_n(N_0^\times + n_0 + n_1) + \alpha(p_0 + p_1)}\Delta p. \qquad \ldots\ldots(21)$$

Substituting this value for Δn in equation (18) and using equation (13)

$$\tau_p = \frac{\alpha_n^{-1}(p_0 + p_1) + \alpha^{-1}(n_0 + n_1 + N_0^\times)}{N_0^\times N_0^- + (n_0 + n_1)N_0^- + (p_0 + p_1)N_0^\times}. \qquad \ldots\ldots(22)$$

Now from equations (6) to (8)

$$\left.\begin{aligned} N_0^\times &= N(1 + n_0/n_1)^{-1} = N(1 + p_1/p_0)^{-1} \\ N_0^- &= N(1 + n_1/n_0)^{-1} = N(1 + p_0/p_1)^{-1}. \end{aligned}\right\} \qquad \ldots\ldots(23)$$

These values (23) substituted in equation (22) lead to equation (1).

3.2. *Initial Lifetime*

It can be seen from equation (16) that in general the surplus recombination of holes depends not only on Δp but also on the change $\Delta N^\times$ of neutral traps. If surplus holes are injected into an n-type semiconductor the lifetime of holes before steady state conditions are obtained is not equal in value to τ_p of equation (1).

Instead as injection is commenced the initial lifetime from equations (13), (16) and (23) is given by

$$\tau_{1n} = \frac{1}{\alpha N_0^-} = \frac{(1 + n_1/n_0)}{\alpha N}. \qquad \ldots\ldots(24)$$

As N^- changes in concentration, so τ increases from τ_{1n} to τ_p.

§ 4. CONCLUSIONS

Hole lifetime in an n-type semiconductor with acceptor traps has been derived by considering recombination and generation processes. A similar analysis may equally well be applied to donor traps, also to electron lifetime in p-type semiconductors.

ACKNOWLEDGMENT

We are greatly indebted to Dr. P. A. Lee for many helpful discussions.

REFERENCES

RANSOM, P., and ROSE, F. W. G., 1954, *Proc. Phys. Soc.* B, **67**, 646.

SCHOTTKY, W., 1952, *Z. Phys.*, **132**, 261.

SCHOTTKY, W., and SPENKE, E., 1939, *Wiss. Veröff. Siemens*, **18**, 225.

SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (London : D. van Nostrand Co.).

SHOCKLEY, W., and READ, W. T., 1952, *Phys. Rev.*, **87**, 835.

VALDES, L. B., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 1420.

Measurements of Current Noise in Lead Sulphide at Audio Frequencies

By D. BARBER
Radar Research Establishment, Great Malvern, Worcs.

Communicated by G. G. Macfarlane; MS. received 18*th April* 1955 *and in amended form* 9*th June* 1955

Abstract. Measurements have been made of the electrical noise produced by passing direct current through lead sulphide photoconductive cells. The mean square noise voltage per unit bandwidth, $\overline{v_n^2}$, was measured in relation to frequency, mean exciting current, cell temperature, and illumination. $\overline{v_n^2}$ was found to vary as the square of the exciting current and to increase with decreasing frequency in a manner which varied from cell to cell. $\overline{v_n^2}$ also increased as the resistance was increased, either by altering the temperature or the illumination, and was dependent on the temperature and illumination only in so far as they affected the resistance.

§ 1. INTRODUCTION

When a direct current is passed through a semiconductor, noise is generated which is in excess of the Johnson noise. Unlike Johnson noise, this excess noise, or current noise as it is called, is frequency dependent, increasing rapidly with decreasing frequency. A number of theories (Macfarlane 1947, 1950, Richardson 1950, van der Ziel 1950, du Pré 1950, Surdin 1951) have been developed in an attempt to derive an expression for the magnitude of current noise.

The results given in this paper are confined to measurements made on one particular form of semiconductor, namely the lead sulphide photoconductive cell of the general form shown in figure 1. The mean square noise voltage per unit bandwidth has been measured in relation to frequency, mean exciting current, temperature of semiconductor and intensity of illumination. The variation of current noise with frequency was investigated over the range 20 c/s to 20 kc/s, and the other measurements were made at chosen frequencies within this range.

§ 2. DESCRIPTION OF APPARATUS

A block diagram of the apparatus is shown in figure 2. The current source which fed the semiconductor consisted of h.t. batteries housed in a screened box, with means for varying and measuring the current through the semiconductor. The semiconductor mount contained h.t. decoupling and other components in addition to the semiconductor. The noise was fed into a battery-operated voltage amplifier with a maximum gain of 70 dB, and thence to a Muirhead–Pametrada wave analyser which selected a narrow band of frequencies, the width of which was a constant small percentage of the centre frequency. The wave analyser output was taken to the primary of a thermocouple via an impedance matching circuit. Since the fluctuating output of the thermocouple was smoothed only by its thermal time constant, it was amplified by a d.c. amplifier and smoothed electronically by a circuit having a time constant of up to 50 seconds. A reading representing the mean square noise voltage within a known bandwidth was obtained on the output meter.

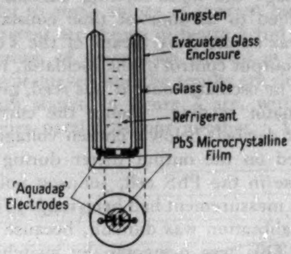

Tungsten
Evacuated Glass Enclosure
Glass Tube
Refrigerant
PbS Microcrystalline Film
'Aquadag' Electrodes

Figure 1. Form of lead sulphide photoconductive cell.

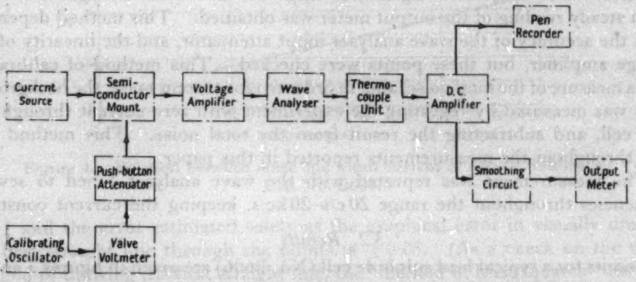

Current Source
Semi-conductor Mount
Voltage Amplifier
Wave Analyser
Thermo-couple Unit
D.C. Amplifier
Pen Recorder
Push-button Attenuator
Smoothing Circuit
Output Meter
Calibrating Oscillator
Valve Voltmeter

Figure 2. Block schematic diagram of noise-measuring apparatus.

Provision was made for the production of a pen-recording of the noise prior to the final smoothing. Calibration was achieved by feeding a measured sinusoidal signal from an oscillator to a small resistance in series with the semiconductor, the procedure being described below.

The entire apparatus was housed in a double-screened cage, and precautions were taken to eliminate both radio-frequency and 50 c/s pick-up.

Typical PbS photoconductive cell dimensions and properties were as follows:

Length of film	0·1–1 cm
Width of film	~1 cm
Thickness of film	$\sim 1\,\mu$
Carrier density (room temperature, dark)	10^{15}–10^{16} cm^{-3}
Mobility (room temperature, dark)	1–10 $\mathrm{cm\ sec^{-1}/v\ cm^{-1}}$.

§ 3. VARIATION OF NOISE WITH FREQUENCY AND CURRENT

3.1. *Experimental Procedure*

These measurements were carried out with the PbS cell at room temperature and in the dark. The wave analyser was used in the condition which gives a bandwidth of 4·25% of the centre frequency. The current flowing in the PbS cell was adjusted to the desired value, and the frequency selected on the wave analyser. After a time equal to several time constants of the smoothing circuit, the reading of the output meter was taken. To express this reading in terms of mean square noise voltage the noise-exciting current was switched off, the smoothing circuit switched to its shortest time constant, and the calibration signal applied. This was adjusted by means of the 1 dB step attenuator and, for fine adjustment, the output control of the oscillator, to give the same reading on the output meter. The oscillator output was read, giving a fractional decibel correction to the attenuator reading. Hence the current noise contribution was obtained in terms of decibels below a chosen voltage.

The reading obtained on the output meter during noise measurement is made up of current noise in the PbS cell, Johnson noise and amplifier noise. It was found that when a measurement had been made in which the current noise component was small, calibration was difficult, because the output meter gave a fluctuating reading. This was overcome by switching the internal input attenuator of the wave analyser, during calibration, so that it was 40 dB less sensitive. The calibration signal was then well above the background noise, and a steady reading of the output meter was obtained. This method depended upon the accuracy of the wave analyser input attenuator, and the linearity of the voltage amplifier, but these points were checked. This method of calibration gave a measure of the *total* noise, and, in order to plot current noise, the background noise was measured by repeating the experiment with zero current through the PbS cell, and subtracting the result from the total noise. This method was used throughout the measurements reported in this paper.

The measurement was repeated with the wave analyser tuned to several frequencies throughout the range 20 c/s–20 kc/s, keeping the current constant.

3.2. *Results*

Results for a typical lead sulphide cell (No. B606) are given in figures 3 and 4. Noise results were obtained in terms of decibels below the chosen calibrating signal, but for clarity in plotting, an arbitrary zero has been chosen and the noise values appear in terms of decibels above this value. The noise spectrum was measured at four values of current 25, 50, 75 and 100 μA, and the background noise at zero current. In each case the total noise values have been corrected for

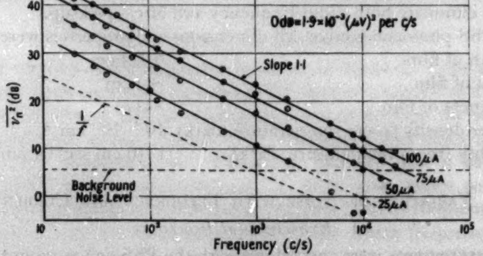


Figure 3. Relation between noise and frequency at various values of mean current. Cell No. B606.

background noise, so the points plotted represent current noise alone. The accuracy of these corrections decreases as the difference between total noise and background noise becomes smaller, which accounts for the increased scatter of the points corresponding to small values of current noise. The noise spectra for the four values of current are shown in figure 3 and the $1/f$ line is drawn for comparison. The variation of noise with current at chosen frequencies was read off from this family of curves, and the results are shown in figure 4. It will be seen that the spectra shown in figure 3 are, within the accuracy of the experiments, straight lines parallel to one another. The slope of the lines is $-1{\cdot}1$ and the error estimated solely as the graphical error in visually drawing the best straight line through the points is $\pm 0{\cdot}05$. (As a check on the visual method of drawing the best straight line, the 'method of least squares' was used for the $100\,\mu\text{A}$ line, and this gave a slope of $-1{\cdot}14$.) This will be the dominant error since each graph contains about fourteen points, the error on each being ± 1 dB. The noise–current curves shown in figure 4 also give parallel straight lines, their slopes being 2·02 with an estimated graphical error of $\pm 0{\cdot}04$.

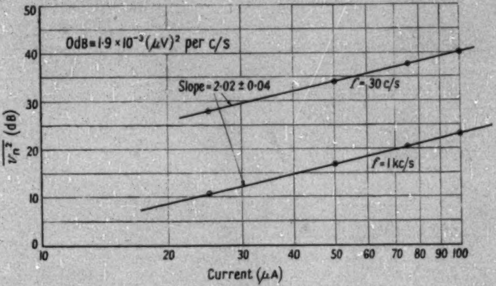


Figure 4. Relation between noise and mean current at chosen frequencies. Cell No. B606.

Not all the cells gave noise spectra which were straight lines of slope greater than unity, when plotted in this manner. The family of curves for one cell (No. C329) whose departure from this was particularly marked, is shown in figure 5. The points have been corrected for background noise, which corresponded to 22 dB, which again accounts for the inaccuracies in the lowest curve. The variation of noise with current for this cell was similar to that shown in figure 4 and gave slopes of $2{\cdot}1 \pm 0{\cdot}1$ and $2{\cdot}0 \pm 0{\cdot}25$ where the errors quoted are the estimated graphical errors. The deviation of the points from the best straight line was greater than in figure 4, especially for $f = 1$ kc/s.

§ 4. VARIATION OF NOISE WITH TEMPERATURE AND ILLUMINATION

4.1. *Introduction*

Assuming the current and measuring frequency to be kept constant, the current noise may still be dependent upon the resistance R of the photoconductive cell, its temperature T, and the illumination L falling upon it, i.e. the mean square noise voltage per unit bandwidth, $\overline{v_n^2} = \phi[R(T, L), T, L]$. For one particular cell it is not possible to vary the resistance without varying one of the other parameters, except by taking various tappings along the film which is impracticable with the usual vacuum cell. No attempt was made to correlate the noise from a number of different cells with their resistances, keeping the temperature and illumination fixed, because this would have introduced a number of other variables, such as electrode separation, purity of semiconductor, crystallite size and film thickness. It was decided to vary the temperature and the illumination separately, each time keeping the other parameter constant, and plot $\overline{v_n^2}$ as a function of cell resistance, in order to compare the curves obtained.

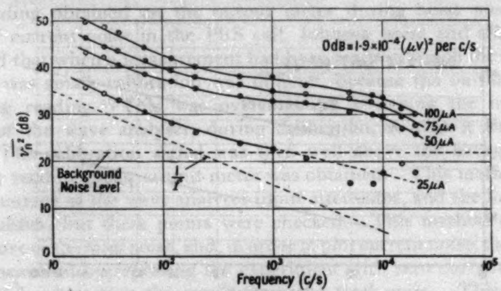


Figure 5. Relation between noise and frequency at various values of mean current. Cell No. C329.

By cooling a lead sulphide cell, a considerable range of resistance can be obtained. For the purpose of these experiments there is, however, a limit to the range of resistance obtainable by varying the illumination at a fixed temperature. Since the refrigerant which keeps the temperature constant is separated from the lead sulphide layer by the glass upon which the layer is deposited, luminous energy falling upon the lead sulphide would cause a temperature gradient across the glass. To avoid this error in the temperature of the lead sulphide, the maximum illumination used was limited to that at which the temperature error was less than the other experimental errors. The temperature of the lead sulphide was then taken as that of the refrigerant. In order to produce graphs of $(\overline{v_n^2}, R(T))$ for constant L and $(\overline{v_n^2}, R(L))$ for constant T which could be compared over a considerable range, the procedure adopted was to use a number of refrigerants giving a wide range of fixed temperatures, and to measure the noise for various values of illumination at each of these. The first point of each illumination curve, corresponding to zero illumination, gave the graph $(\overline{v_n^2}, R(T))$ for constant L, covering a wide range of R, with which the illumination curves, each covering a smaller range, could be compared.

4.2. *Experimental Procedure*

The refrigerants used were ice, methyl chloride, Freon 12, solid carbon dioxide and liquid ethylene. Solid carbon tetrachloride and solid methyl alcohol were also used but only under zero illumination. The $(\overline{v_n^2}, R(T))$ graph was extended above room temperature by placing a previously heated brass cylinder in the cell receptacle, and making noise measurements as it gradually cooled down. The ($\overline{v_n^2}$, $R(L)$) graph for T fixed at room temperature was also obtained.

When the cells were cooled, trouble was caused by condensation of atmospheric water vapour on the glass around the tungsten leads. To avoid this condensation leakage the cell leads were lengthened by spot-welding nickel wires to them, and sliding polythene sleeving over the weld and an inch or so of the nickel wire. Any gap between the polythene and the glass pinch was sealed with paraffin wax or picein. The top part of the polythene sleeve was free from condensation, so leakage was avoided.

Once the cell was prepared it was mounted in a re-designed cell head which was fitted with a light-proof shutter above the cell for replenishment of refrigerant. The illumination was provided by a 24-v 36-w bulb focused by a lens through a second shutter in the side of the cell mount and it was deflected on to the lead sulphide film by a 45° mirror placed below the cell.

The cell receptacle was filled with refrigerant and both shutters were closed. Arrangement was made to measure the resistance of the cell *in situ*. As the temperature of the lead sulphide fell, its resistance increased. When this reached a maximum it was assumed that the lead sulphide had acquired the temperature of the refrigerant. The cell current and measuring frequency were set to fixed values, the noise measurement and calibration carried out in the way described in §3.1 and the resistance measurement repeated. If there were any slight changes in resistance due to temperature drift, the mean of the two readings was taken.

This measurement gave the noise corresponding to a certain temperature and resistance, under zero illumination. The illumination was adjusted to the required value, and the resistance measurements, noise measurement and calibration were repeated. This was carried out for a number of values of illumination to obtain the illumination curve relevant to that temperature.

Similar experiments were performed with the other refrigerants mentioned above. In the case of solid refrigerants, particularly carbon dioxide, it was sometimes necessary to press the refrigerant down into the cell receptacle, between noise measurements, to maintain good thermal contact.

It was found possible to obtain zero illumination noise measurements at temperatures between the refrigerant temperatures by using a solid refrigerant, allowing it to melt, and making a noise measurement and calibration as rapidly as possible, as the cell temperature rose. The mean of the resistance values before and after noise measurement was taken. Some accuracy was lost, but this formed a useful check on the shape of the ($\overline{v_n^2}$, $R(T)$) curve where the temperature gap between refrigerants was large. For example, methyl alcohol was used in this way to obtain points between those corresponding to Freon 12 and carbon dioxide.

An additional correction had to be applied to the figure representing the mean square noise voltage per unit bandwidth, because the cell was not being fed from a constant current source. If the noise is regarded as a fluctuation in the resistance of the cell, then the voltage fluctuation caused by passing a constant current I through the cell will be $\overline{v_n^2} = I^2\overline{\delta R^2}$. If, however, the load resistance through which the cell is fed is comparable with the cell resistance, the current will not be constant. Any increase in the cell resistance, which would cause a positive voltage signal, also causes the cell current to fall, which degenerates the voltage signal. This degeneration of the noise depends on the value of the cell resistance, and since this was being varied in these experiments, it was necessary to correct for it. It can be shown that this is achieved by multiplying the measured mean square noise voltage by $[(R_L + R)/R_L]^2$ where R_L is the load resistance and R the cell resistance.

4.3. *Results*

The results for a typical cell (No. 3) are shown in figures 6 and 7. Figure 6 shows the graph of mean square noise voltage per unit bandwidth plotted against log resistance for various temperatures under zero illumination, i.e. $(\log \overline{v_n^2}, \log R(T))$. A smooth curve has been drawn through the points, and this curve has also been drawn in for comparison on figure 7. The slope of the curve increases from 1·4 to 4·7 as R increases. The points plotted in figure 7 show the variation of noise with illumination, corresponding to the various fixed temperatures. It can be seen that the points on the illumination

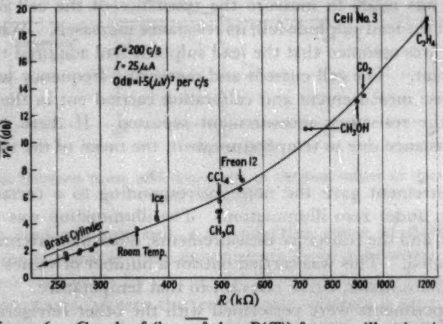


Figure 6. Graph of $(\log \overline{v_n^2}, \log R(T))$ for zero illumination.

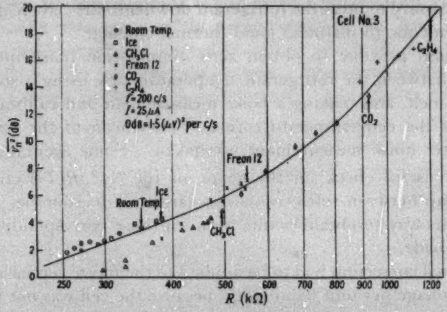


Figure 7. Graph of $(\log \overline{v_n^2}, \log R(L))$ for a number of fixed temperatures.

curves lie within about ± 2 dB of the temperature curve. The illumination curves follow the shape of the temperature curve rather better than this, for the starting point of each illumination run—corresponding to zero illumination—may not lie exactly on the temperature curve. This is particularly true of the methyl chloride illumination curve. The probable error on any one reading includes the error in the noise reading and that of the resistance measurement, and is estimated as ± 1·5 dB. It appears that, within the accuracy of the experiments, the noise delivered by a given cell is dependent only upon its resistance, and hence is dependent on the temperature and illumination only because they affect the resistance. The mean square noise voltage per unit bandwidth increases as the cell resistance is increased.

The fact that the dark resistance of the lead sulphide cells was measured at each temperature, enabled them to be used as their own thermometers. Accordingly a separate experiment was carried out to obtain the dark-resistance–temperature characteristic of each cell, figure 8. It was then possible to plot the mean square noise voltage per unit bandwidth against temperature (for zero illumination) instead of against resistance as in figure 6. Such a plot showed that $\overline{v_n^2}$ increased as the temperature was reduced.

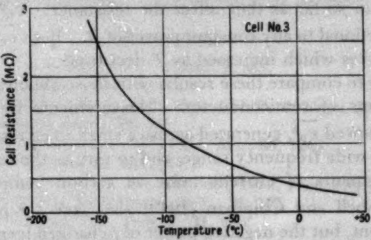


Figure 8. Variation of resistance of PbS cell with temperature.

The corresponding values of mean square noise voltage per unit bandwidth $\overline{v_n^2}$, resistance R, and absolute temperature T, were then known. Since noise *power* may be more fundamental than $\overline{v_n^2}$, the values of $\overline{v_n^2}/R$ have been worked out and plotted against T, as shown in figure 9. As T is reduced, the noise power still increases, though, of course, less rapidly than $\overline{v_n^2}$. This is in contrast to the temperature dependence of Johnson noise in a wire-wound resistor, for which $\overline{v_n^2}/R$ varies linearly with T.

Bell (1955) has pointed out that the present results give a linear relation between $\log \overline{v_n^2}$ and $1/T$.

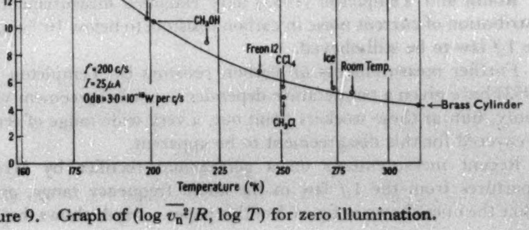


Figure 9. Graph of ($\log \overline{v_n^2}/R$, $\log T$) for zero illumination.

§ 5. CONCLUSION

The variation of mean square noise voltage per unit bandwidth $\overline{v_n^2}$ with frequency, current, temperature and illumination has been investigated.

For all the eight measured cells, $\overline{v_n^2}$ increased with decreasing frequency over the range 20 c/s to 20 kc/s. For some cells, the value of y in $\overline{v_n^2} \propto f^{-y}$ was constant and close to unity, whereas for others it differed widely from unity and varied on either side of unity over the above frequency range. In all the cells measured, $\overline{v_n^2}$ was proportional to the square of the exciting current, within the experimental accuracy.

For fixed frequency and current, the graphs of $(\overline{v_n^2}, R(T))$ for zero illumination and $(\overline{v_n^2}, R(L))$ for fixed temperatures coincided, showing that the noise was dependent only upon the resistance, and was dependent on the temperature and illumination only in so far as they affect the resistance. $\overline{v_n^2}$ increased with R, but was not proportional to any constant power of R. Both $\overline{v_n^2}$ and $\overline{v_n^2}/R$ plotted against T gave curves which increased as T decreased.

It is interesting to compare these results with those obtained by other workers using various forms of semiconductor. Measurements by Miller (1947) on crystal rectifiers showed $\overline{v_n^2}$, generated across a small fixed resistor, to be proportional to $1/f$ over a wide frequency range, and to vary as the square of the reverse current. Measurements of current noise in carbon composition and other resistors by Campbell and Chipman (1949) also gave $\overline{v_n^2}$ proportional to the square of the current, but the negative power of f changed from less than unity to 1·6 over the frequency range 20 kc/s to 500 kc/s.

Harris, Abson and Roberts (unpublished, see also Harris 1948) found that for lead sulphide films and carbon resistors, the index of f was in the region of -1, and the index of current around 2, but no estimate of experimental accuracy was made. They also found the noise to increase at low temperatures.

In a review article, however, Angello (1949) quotes further results of Miller as showing that the temperature variation of current noise in semiconductor rectifiers is not large.

Departures from the $1/f$ law at frequencies depending upon the time constants of the physical processes involved have been observed in germanium single crystals by Herzog and van der Ziel (1951) and by Montgomery (1952). The variation of noise power with temperature reported in the present paper is more regular than that found by Montgomery for germanium single crystals, but the measurements do not cover such a wide temperature range.

Rollin and Templeton (1953) have extended measurements of the spectral distribution of current noise in carbon resistors to below 10^{-3} c/s, and have shown the $1/f$ law to be still obeyed.

Further measurements on carbon resistors by Templeton and MacDonald (1953) have given a temperature dependence not in agreement with Macfarlane's theory, but, as these workers point out, a very wide range of temperatures must be covered for this disagreement to be apparent.

Recent measurements on a germanium rectifier by Hyde (1953) show departures from the $1/f$ law in the audio frequency range, giving spectra not unlike the ones shown in figure 5 of this paper. Hyde shows that his experimental curves can be synthesized by the superposition of terms of the form

$$Af^{-1},\quad B[1+(2\pi f\tau_1)^2]^{-1},\ C[1+(2\pi f\tau_2)^2]^{-1}$$

where τ_1, τ_2 are relaxation times of the rectifier.

The results for lead sulphide given in the present paper are consistent with the bulk of the reported measurements for similar forms of semiconductor. The further evidence, and particularly that on the temperature dependence of the noise power and the effect of illumination, may help to check the predictions of the several theories on current noise.

ACKNOWLEDGMENTS

The author wishes to thank Professor J. R. Whitehead for advice in the early stages of this work, and several colleagues for helpful discussions. Acknowledgment is made to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

ANGELLO, S. J., 1949, *Elect. Engng, N.Y.*, October, p. 865.
BELL, D. A., 1955, *Brit. J. Appl. Phys.*, **6**, 284.
CAMPBELL, R. H., and CHIPMAN, R. A., 1949, *Proc. Inst. Radio Engrs, N.Y.*, **37**, 938.
HARRIS, E. J., 1948, *Electronic Engng*, **20**, 145.
HERZOG, G. B., and VAN DER ZIEL, A., 1951, *Phys. Rev.*, **84**, 1249.
HYDE, F. J., 1953, *Proc. Phys. Soc.* B, **66**, 1017.
MACFARLANE, G. G., 1947, *Proc. Phys. Soc.*, **59**, 366; 1950, *Proc. Phys. Soc.* B, **63**, 807.
MILLER, P. H., 1947, *Proc. Inst. Radio Engrs, N.Y.*, **35**, 252.
MONTGOMERY, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
DU PRÉ, F. K., 1950, *Phys. Rev.*, **78**, 615.
RICHARDSON, J. M., 1950, *Bell Syst. Tech. J.*, **29**, 117.
ROLLIN, B. V., and TEMPLETON, I. M., 1953, *Proc. Phys. Soc.* B, **66**, 259.
SURDIN, M., 1951, *J. Phys. Radium*, **12**, 777.
TEMPLETON, I. M., and MACDONALD, D. K. C., 1953, *Proc. Phys. Soc.* B, **66**, 680.
VAN DER ZIEL, A., 1950, *Physica*, **16**, 359.

The Turnover Phenomenon in Thermistors and in Point-Contact Germanium Rectifiers

By R. E. BURGESS†
Department of Scientific and Industrial Research, Radio Research Station, Slough

Communicated by R. L. Smith-Rose; MS. received 16*th September* 1954 *and in revised form* 13*th June* 1955

Abstract. Thermistors are semiconducting devices which display a voltage maximum in their steady-state current voltage characteristics. The properties of this 'turnover' phenomenon can be calculated from the dependence of the thermistor conductance on the temperature of the element and for an exponential relation of the form $I \propto V \exp(-b/T)$ it is found that the turnover power increases as the square of the absolute ambient temperature.

The similarity in form between the reverse characteristic of a point-contact germanium rectifier and the static characteristic of a thermistor suggests a thermal hypothesis for rectifier turnover. However, in the latter, the turnover power decreases approximately linearly with the ambient temperature. Even if it be assumed that the isothermal characteristic is not linear and that the energy term b is a function of applied voltage such as would be expected in a number of physical models, the theoretical turnover power increases with increasing ambient temperature.

A number of possible forms for the dependence of rectifier current on the voltage and contact temperature are examined and a general type of function which displays the essential features of the observed turnover behaviour is discussed. The physical basis for such a dependence does not conform with existing theories of rectification and a radical revision of these is apparently required under the conditions of elevated contact temperature and high field which exist at turnover.

§ 1. Introduction

The advent of the point-contact germanium rectifier (Benzer 1949) capable of withstanding large reverse voltages (100 volts or more) has led to speculation concerning the nature of the 'turnover' phenomenon which limits the reverse voltage which can be applied to such a rectifier. It is known that the turnover voltage (i.e. the voltage at which the slope of the static current–voltage characteristic becomes infinite) is a function of ambient temperature and of the electrical forming treatment which has been applied to give a high reverse resistance.

The apparent similarity between the static reverse characteristic of a germanium rectifier and the static characteristic of a thermistor has led to the suggestion that turnover is essentially a thermal phenomenon and additional experimental evidence gives some support to this contention. Briefly it is agreed that as the reverse current through the rectifier increases the contact temperature rises, so lowering the contact resistance until the latter decreases at such a rate that the voltage ceases to increase with increasing current. In this case the condition of turnover in a rectifier whose current I can be expressed as a function of voltage V and contact temperature T:

---

† Now at the Department of Physics, University of British Columbia, Vancouver.

$$I = f(V, T)$$

will be given by

$$1 = V_1 \left(\frac{\partial I}{\partial T}\right)_{\substack{V=V_1\\T=T_1}} \left(\frac{dT}{dP}\right)$$

where V_1 is the turnover voltage, T_1 is the temperature of the contact at turnover and P is the electrical power IV. If Newton's law of cooling is valid, (dT/dP) will be a constant.

In a thermistor (Becker *et al.* 1947, Burgess 1952) where field effects are unimportant there is close adherence of the static characteristic to that predicted on a purely thermal basis with Newtonian cooling and the isothermal current–voltage characteristic is ohmic.

In the germanium rectifier, however, the isothermal reverse characteristic is not linear and strong field effects exist, and it would not therefore be surprising if turnover appeared as a combination of field and temperature effects. The purpose of the present paper is first to show that an explanation of turnover, using the arguments applied to the thermistor, is untenable since in a thermistor the turnover power increases with increasing ambient temperature whereas in a germanium rectifier it decreases. It is further demonstrated that other current–voltage characteristics which are plausible on physical grounds lead to the same erroneous conclusion. On the other hand, characteristics which do predict a turnover behaviour in accordance with experimental observation are not of the form which would be anticipated in terms of the customary physical models of rectifying barriers. Thus the problem of devising a satisfactory physical model of turnover still remains.

§ 2. TURNOVER IN A THERMISTOR

The steady-state characteristic of a thermistor is known to display two extrema of voltage, first a maximum V_1 then, as the current further increases, a minimum V_2 (figure 1). This static characteristic (which refers to some particular ambient temperature) is the relation between current and voltage when the temperature of the thermistor is allowed to attain equilibrium at each point so that the electrical power IV is exactly balanced by the heat loss from the thermistor.

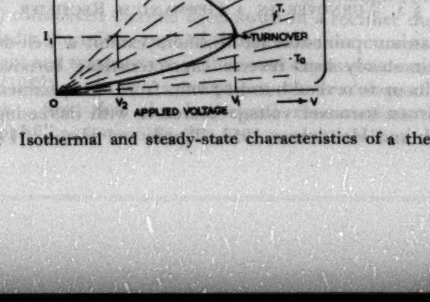

STEADY-STATE CHARACTERISTICS (FOR AMBIENT TEMP. Ta)
I
I₂
CURRENT
INCR. TEMP.
ISOTHERMAL CHARACTERISTICS
T₁
I₁
TURNOVER
Ta
O
V₂
V₁
V
APPLIED VOLTAGE

Figure 1. Isothermal and steady-state characteristics of a thermistor.

Now the voltage maximum V_1 in the steady-state characteristic is called the 'turnover voltage' and for a given thermistor is a function only of the ambient temperature T_a at which the characteristic applies.

The property of turnover arises from the temperature dependence of the conductivity of the semiconducting material of which the thermistor is made. Over the temperature range of interest the current–voltage relation is closely given by

$$I = VC\exp(-b/T) \qquad \ldots\ldots(1)$$

where C and b are constants and T is the absolute temperature. The effective activation energy associated with the conductivity of the material is $\boldsymbol{k}b$ where $\boldsymbol{k}$ is Boltzmann's constant.

If it be further assumed that the heated element at T loses heat to the surroundings at T_a in accordance with Newton's law of cooling, then

$$T - T_a = aP \qquad \ldots\ldots(2)$$

where a is the constant which relates the excess temperature to the electrical power supplied to the element. Elementary calculation now shows that the absolute temperature of the element at turnover is given by

$$T_1 = \tfrac{1}{2}b - [\tfrac{1}{4}b^2 - bT_a]^{1/2} \qquad \ldots\ldots(3)$$

where the suffix 1 will throughout denote the value at turnover. It is seen that unless $b > 4T_a$ turnover will not occur.

The power at turnover is given by

$$P_1 = a^{-1}[\tfrac{1}{2}b - T_a - (\tfrac{1}{4}b^2 - bT_a)^{1/2}] \qquad \ldots\ldots(4)$$

which shows that *the turnover power increases with increasing ambient temperature.* This is the important feature of a thermistor or any semiconducting device whose conductance depends on absolute temperature in the manner of equation (1).

If, as is common to thermistors (or in filaments of relatively pure germanium), b is much greater than T_a, the turnover power is to a close approximation given by

$$I_1V_1 \equiv P_1 \simeq T_a{}^2/ab \qquad \ldots\ldots(5)$$

while the conductance at turnover is $I_1/V_1 \equiv G_1 \simeq G_0 \exp(1 + T_a/b)$ showing that at turnover the thermistor conductance is slightly more than e times its isothermal value G_0 at the same ambient temperature.

In a typical bead-type thermistor $b = 3000$–4000 degrees so that turnover will occur at all normal ambient temperatures. However the power required to attain the voltage minimum at V_2 is so great that it would never be reached in practice since the bead temperature there would be given by $T_2 = \tfrac{1}{2}b + (\tfrac{1}{4}b^2 - bT_a)^{1/2}$ which, being only slightly less than b would be quite unattainable.

§ 3. TURNOVER IN A GERMANIUM RECTIFIER

N-type germanium point-contact rectifiers exhibit a well-defined turnover condition in their steady-state reverse characteristics; turnover voltage may range from 10 volts up to several hundred volts from one specimen to another and for a given specimen turnover voltage decreases with increasing ambient temperature (Tillman and Henderson 1953, Tipple and Henisch 1953) (figure 2).

Benzer (1949) conjectured that turnover was a thermal effect which arose from the rapid decrease of rectifier resistance with increasing contact temperature. In support of this view he pointed out that the power at turnover P_1 was approximately a linearly decreasing function of ambient temperature T_a over an appreciable range of temperature (figure 3). The intercept of the linear region of the (P_1, T_a) plot with the T_a axis was interpreted as a certain critical temperature T_c to which the contact is brought in the turnover condition, whatever the ambient temperature T_a; furthermore the slope of the plot was related to the power required to produce unit rise of contact temperature. Benzer conjectured that T_c was in the range of temperature corresponding to the onset of intrinsic conductivity in the germanium but this statement was not elaborated.

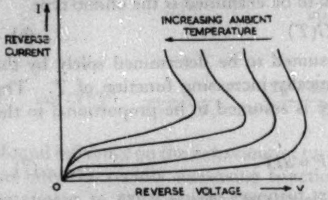


Figure 2. Turnover in the steady-state reverse characteristic of a germanium point contact rectifier.

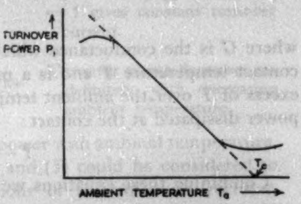


Figure 3. Observed dependence of turnover power in a germanium rectifier on ambient temperature.

Hunter (1951) also subscribed to the self-heating view and attempted to synthesize the steady-state characteristic on the assumption that the isothermal characteristics were straight lines corresponding to a conductance which increased rapidly with increasing temperature; some confirmation of the linearity of the isothermal characteristics was obtained by Bennett and Hunter (1951) who used $\frac{1}{2}$-microsecond pulses to eliminate heating effects. The intercepts of these isothermals with the hyperbolae of constant power were then assumed (on the basis of Newton's law of cooling) to map out the static characteristics. This is precisely the procedure applicable to a thermistor and therefore if a temperature dependence of contact conductance of the form $\exp(-b/T)$ be assumed although the static thermistor and germanium rectifier characteristics exhibit marked qualitative similarity (Burgess 1952) they differ completely in the essential feature that in a thermistor the turnover power increases with increasing ambient temperature, while in the germanium rectifier it decreases. An attempt is now made to examine quantitatively the phenomenon of turnover in a germanium rectifier, and it will be seen that certain conditions are required in order that turnover may occur at all and that further more stringent conditions are needed if the turnover power is to decrease with increasing ambient temperature.

Billig (1951) considered thermal breakdown in a rectifier due to heating and effectively applied a thermistor type of theory since the current was determined by a barrier height and temperature in the usual exponential manner.

Armstrong (1953) made two assumptions which elaborate the model used in earlier work: first that the isothermal characteristic is of the form $A\exp[(bV^{1/2}-c)/\boldsymbol{k}T]$ and secondly that heat conduction from the contact occurs by way of the germanium whose thermal conductivity varies inversely as the absolute temperature but the outcome of these hypotheses is that the turnover power increases with the ambient temperature.

Consideration will now be given to the dependence of the turnover current, voltage and power on the ambient temperature for various forms of functional dependence of the current I on voltage V and contact temperature T. Since we are only concerned with phenomena occurring in the voltage range above about 10 volts, the assumed forms of characteristic are those applicable in this range and do not relate to the form of the current for smaller voltages.

§ 4. LINEAR CHARACTERISTICS

The first type of current characteristic to be examined is the ohmic type

$$I = VG(T) \qquad \ldots\ldots(6)$$

where G is the conductance which is assumed to be determined solely by the contact temperature T and is a monotonically increasing function of T. The excess of T over the ambient temperature is assumed to be proportional to the power dissipated at the contact:

$$T - T_a = aP = aIV. \qquad \ldots\ldots(7)$$

Combining these equations we see that turnover occurs when

$$T_1 - T_a = G(T_1)/G'(T_1) \qquad \ldots\ldots(8)$$

which determines T_1 and hence I_1, V_1 and P_1. Turnover can only occur when this equation has at least one real root. The rate of change of turnover voltage V_1 with ambient temperature T_a is given by

$$\frac{dV_1}{dT_a} = -\frac{1}{2aI_1} \qquad \ldots\ldots(9)$$

which is *always* negative. Thus the condition that the turnover voltage shall decrease with increasing ambient temperature is no criterion of acceptability for the form of $G(T)$. On the other hand the rate of change of turnover power with ambient temperature is given by

$$dP_1/dT_a = (r-1)/a \qquad \ldots\ldots(10)$$

where the parameter r is given by

$$r = dT_1/dT_a = [G'(T_1)]^2/G(T_1)G''(T_1). \qquad \ldots\ldots(11)$$

In the case of a thermistor $r > 1$ but it is required in a germanium rectifier that $r < 1$. The rate of change of turnover current with T_a may also be expressed in terms of the parameter r:

$$dI_1/dT_a = (2r-1)/2aV_1 \qquad \ldots\ldots(12)$$

showing that, according as r is greater or less than half, the turnover current will increase or decrease with increasing ambient temperature.

It is now appropriate to examine a number of analytical forms of $G(T)$ and to determine in particular P_1 and r for each case. This is done in the table.

Figure 4 illustrates the dependence of turnover power on ambient temperature for the cases presented in the table. The most striking feature is that although all the functions $G(T)$ are increasing functions of T their exact form exerts a

Turnover Parameters for various Forms of Linear Isothermal Characteristics

Form	$G(T)\propto$	r	P_1	Remarks
1	$(1+T/b)^n$	$n/(n-1)$	$\dfrac{T_\mathrm{a}+b}{a(n-1)}$	$n>1$ for turnover to occur.
2	$\exp(T/b)$	1	b/a	Constant turnover power.
3	$\exp(-b/T)$	$\dfrac{1}{1-2T_1/b}$	$\dfrac{b-2T_\mathrm{a}-(b^2-4bT_\mathrm{a})^{1/2}}{2a}$	Thermistor characteristic $b>4T_\mathrm{a}$ for turnover to occur.
4	$(1-T/b)^{-n}$	$\dfrac{n}{n+1}$	$\dfrac{b-T_\mathrm{a}}{a(n+1)}$	Linear decrease of turnover power with ambient temperature up to $T_\mathrm{a}=b$. $n=1$ gives constant turnover current.
5	$\exp(T^2/b^2)$	$\dfrac{1}{1+b^2/2T_1^2}$	$\dfrac{(T_\mathrm{a}^2+2b^2)^{1/2}-T_\mathrm{a}}{2a}$	Turnover power decreases indefinitely with increasing ambient temperature.

profound influence on the behaviour of turnover power with ambient temperature. Thus only the rapidly increasing functions (4) and (5) could be considered to approximate to the observed turnover behaviour of a germanium rectifier.

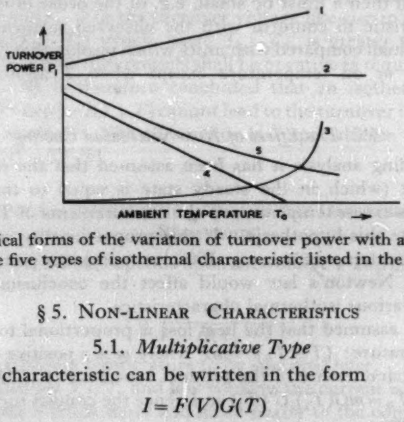


Figure 4. Theoretical forms of the variation of turnover power with ambient temperature for the five types of isothermal characteristic listed in the table.

§ 5. NON-LINEAR CHARACTERISTICS

5.1. *Multiplicative Type*

If the current characteristic can be written in the form

$$I=F(V)G(T) \qquad \ldots\ldots(13)$$

then the current–voltage characteristic will have the same shape at all temperatures, and the function $G(T)$ serves as a scaling factor.

The turnover relation for temperature is exactly the same as for the linear isothermal $VG(T)$, that is to say it is independent of the form of $F(V)$:

$$T_1-T_\mathrm{a}=G(T_1)/G'(T_1)$$

and consequently the turnover temperature coefficient is as before

$$r\equiv\frac{dT_1}{dT_\mathrm{a}}=\frac{[G'(T_1)]^2}{G(T_1)G''(T_1)}.$$

The rates of change of turnover power and turnover voltage with ambient temperature are then

$$\left.\begin{aligned} \frac{dP_1}{dT_a} &= \frac{r-1}{a} \\ \frac{dV_1}{dT_a} &= -\frac{1}{aI_1}\,\frac{F(V_1)}{F(V_1)+V_1F'(V_1)} \end{aligned}\right\} \qquad \ldots\ldots(14)$$

The requirement that dP_1/dT_a shall be negative is that $(d^2/dT^2)[\ln G(T)]>0$.

We now examine the characteristic

$$I=AV^m/(b-T)^n \qquad \ldots\ldots(15)$$

and it is found that the turnover relations are

$$\left.\begin{aligned} &T_1=(nT_a+b)/(n+1) && P_1=(b-T_a)/a(n+1) \\ &V_1=\frac{n(b-T_a)^{(n+1)/(m+1)}}{n+1}(anA)^{1/(m+1)}, && I_1=A\,\frac{n(b-T_a)^{(m-n)/(m+1)}}{n+1}(anA)^{m/(m+1)} \end{aligned}\right\} \qquad \ldots\ldots(16)$$

The turnover power decreases linearly with increasing ambient temperature. The behaviour of the turnover current depends upon the sign of $m-n$ and it remains constant if $m=n$. The rate of change of turnover contact temperature with ambient temperature is $n/(n+1)$ and if, as experiment seems to indicate, this is to be small then n must be small, e.g. of the order of 0·1. Thus for this type of characteristic to conform with the observed turnover behaviour, both m and n must be small compared with unity which implies that the current depends little on voltage or on temperature except when T approaches the critical temperature b.

5.2. *The Effect of Non-Newtonian Cooling*

In the preceding analysis it has been assumed that the rate of loss of heat from the contact (which in the steady state is equal to the input power) is proportional to its excess temperature; the measurements of Tipple and Henisch (1953) corroborate this hypothesis. If this proportionality is not valid (say for large contact excess temperature, e.g. 100°c or more) it is pertinent to examine if departures from Newton's law would affect the conclusions concerning the acceptability of various isothermal characteristics.

First let it be assumed that the heat loss is proportional to the mth power of the excess temperature: $(T-T_a)^m=aP$ where m is a positive constant.

Then for a characteristic of the product type $I=F(V)G(T)$ the condition at turnover is $T_1-T_a=mG(T_1)/G'(T_1)$ and hence the contact turnover temperature coefficient becomes

$$r\equiv\frac{dT_1}{dT_a}=\left[1-m\left(1-\frac{G(T_1)G''(T_1)}{G'^2(T_1)}\right)\right]^{-1}=\left[1+m\left(\frac{1}{r_0}-1\right)\right]^{-1} \qquad \ldots\ldots(17)$$

where r_0 is the value of r which obtains for Newtonian cooling. It is readily seen that if the requirement of $r_0<1$ is not met by any particular choice of $G(T)$ in the case of Newtonian cooling ($m=1$) it is still not satisfied for non-Newtonian cooling with any arbitrary value of m, and thus no hitherto unacceptable model is rendered acceptable. If however the choice of $G(T)$ does give $r_0<1$ then the effect of the revised cooling law (that is, of m) is to give a smaller value of r if $m>1$.

§ 6. CURRENT DEPENDING ON A VOLTAGE-DEPENDENT BARRIER OR EXCITATION PROCESS

A very general class of current characteristics suggested by a variety of physical mechanisms is

$$I = AV^n \exp[-f(V)/T] \qquad \ldots\ldots(18)$$

where T is the absolute temperature of the contact. This type of characteristic is indicated by theories of rectification in which the reverse current due to electron flow from the negative whisker to the germanium is increased by the effects of image force and field emission. In general any field-dependent (and hence V-dependent) barrier height or carrier excitation process would lead to a characteristic of this form and it is therefore of particular physical interest.

Now without detailed evaluation of the turnover parameters certain general results about the rate of change of turnover power and turnover voltage with ambient temperature may be derived:

$$\left.\begin{aligned} &\frac{dP_1}{dT_a} = \frac{r-1}{a} \\ \text{and}\quad &\frac{dV_1}{dT_a} = \frac{1}{a(n+1)I_1}\,\frac{rT_1-(2r-1)T_a}{(T_1-T_a)} \end{aligned}\right\} \qquad \ldots\ldots(19)$$

where $r \equiv dT_1/dT_a$ as before.

For the turnover power to decrease with increasing ambient temperature r must be less than unity. For the turnover voltage likewise to decrease r must be greater than $T_a/(2T_a - T_1)$ which, since $T_1 > T_a$, is greater than unity. Thus the requirements that both differentials shall be negative as required by experiment are incompatible. It is therefore concluded that an isothermal characteristic of the form $I = AV^n \exp[-f(V)/T]$ cannot lead to the turnover behaviour observed in a germanium rectifier. The thermistor is a special case of this with $n=1$, $f(V)=b$ and $r=1/(1-2T_1/b)$.

§ 7. FORMS OF CHARACTERISTIC YIELDING THE OBSERVED TURNOVER BEHAVIOUR

Consider a current which depends upon the voltage V and contact temperature T as

$$I = \frac{AV}{V+B(T_0-T)} \qquad \ldots\ldots(20)$$

where T_0 is some arbitrary temperature to be specified later. This characteristic has the property that for $T<T_0$ and low voltages the current is proportional to V while as V increases it tends more and more nearly to the constant value A. If the contact temperature T is equal to T_0 the current is equal to A independently of V, while if $T>T_0$ the current decreases as V increases.

The turnover relations it leads to are:

$$\left.\begin{aligned} V_1 &= \frac{B(T_0-T_1)}{(ABa)^{1/2}-1} = \frac{B(T_0-T_a)}{2(ABa)^{1/2}-1} \\ I_1 &= (A/Ba)^{1/2} \\ T_1 &= \frac{(ABa)^{1/2}T_0 + [(ABa)^{1/2}-1]T_a}{2(ABa)^{1/2}-1}. \end{aligned}\right\} \qquad \ldots\ldots(21)$$

Now if it be postulated that $ABa=1$ these relations become

$$V_1=B(T_0-T_a), \quad I_1=A, \quad P_1=\frac{(T_0-T_a)}{a}, \quad T_1=T_0 \qquad \ldots\ldots(22)$$

which imply constant turnover current, constant turnover contact temperature and values of turnover voltage and power which decrease linearly with ambient temperature giving an intercept of T_0 and a slope of $-1/a$. These four features are precisely those which are observed in the experimental observations of Tipple and Henisch (1953). For the rectifier referred to in their figure 3 the following parameters are applicable:

$$T_0=135°\text{c}, \quad \text{A}=12\,\text{mA}, \quad B=0{\cdot}7\,\text{v deg}^{-1}, \quad 1/a=8{\cdot}6\,\text{mw deg}^{-1}$$

for a voltage range of $V=30$–90 volts. These figures lead to the isothermal characteristics (broken lines) and also to the associated turnover properties.

This type of characteristic has the property that at turnover the contact temperature always comes to T_0 so rendering the current independent of voltage.

The simple characteristic of equation (20) can be elaborated to retain the essential turnover behaviour described by equation (22). For instance if

$$I=A\left[\frac{V}{V+B(T_0-T)}\right]^n \quad \text{and} \quad nABa=1 \qquad \ldots\ldots(23)$$

then $V_1=nB(T_0-T_a)$, $I_1=A$, $P_1=(T_0-T_a)/a$, $T_1=T_0$. Thus independent determination of the parameters A, B and a permit the determination of n from $n=1/ABa$. Experimental evidence points to n being of the order of unity so the simpler form considered above may be reasonably accurate. It will be noted that I in (23) is a function of the composite variable $(T_0-T)/V$ and further examination shows that other functions of it give the required turnover behaviour For instance

$$I=A\exp\left[-B(T_0-T)/V\right] \quad \text{with} \quad ABa=1 \qquad \ldots\ldots(24)$$

leads to the turnover relations given in equation (22).

Thus it is seen that there is a class of current functions which have the property of attaining a constant turnover current corresponding to a constant turnover contact temperature T_1 equal to T_0.

With $x\equiv(T_0-T)/AaV$ the characteristics given by equations (20), (23) and (24) are respectively of the forms

$$\frac{I}{A}\equiv f(x)=\frac{1}{1+x}, \quad \frac{1}{(1+x/n)^n} \quad \text{and} \quad e^{-x}$$

which have the properties $f(0)=1$ and $f'(0)=-1$ and can all be expanded in the neighbourhood of $x=0$ as the series $f(x)=1-x+$ higher powers of x. In the steady-state where $T=T_a+aIV$ the parameter x has the simple interpretation $x=V_1/V-I/I_1$.

Two problems are posed by these characteristics: (i) what combination of effects of temperature and electric field give rise to this form of equation for the current, i.e. to the parameter x; (ii) what is the physical significance of the constant turnover current A and the critical temperature T_0.

Existing rectifier theories give no indication of these features and radical revision is needed to explain the reverse current in the conditions of high electric field and elevated contact temperature which occur for voltages of above about 10 or 20 volts. It is possible that when the contact is at the temperature T_0 (which usually lies in the range 100–200°c) there is virtually no barrier for the predominant current carrier at the contact and the system is then rather similar to that of a non-rectifying material whose electrical conductivity varies radially with the temperature and the field. One influence of the large electric field which may exist near the contact under turnover conditions would be to reduce the carrier mobility; if a barrier were still present, the effect of the reduced mobility would be to tend to make the results of the diffusion and diode theories coincide (Burgess 1953). In the absence of a barrier it would be necessary to regard the conductivity as a function of both the local field E and temperature T.

ACKNOWLEDGMENTS

The work described above was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

ARMSTRONG, H. L., 1953, *J. Appl. Phys.*, **24,** 1332.
BECKER, J. A., GREEN, C. B., and PEARSON, G. L., 1947, *Bell. Syst. Tech. J.*, **26,** 170.
BENNETT, A. I., and HUNTER, L. P., 1951, *Phys. Rev.*, **81,** 152.
BENZER, S., 1949, *J. Appl. Phys.*, **20,** 804.
BILLIG, E., 1951, *Proc. Roy. Soc.* A, **207,** 156.
BURGESS, R. E., 1952, *Research, Lond.*, 1952, **5,** 469; 1953, *Proc. Phys. Soc.* B, **66,** 430.
HUNTER, L. P., 1951, *Phys. Rev.*, **81,** 151.
TILLMAN, J. R., and HENDERSON, J. C., 1953, *Phil. Mag.*, **44,** 677.
TIPPLE, P. M., and HENISCH, H. K., 1953, *Proc. Phys. Soc.* B, **66,** 826.

The Use of a Modulated Light Spot in Semiconductor Measurements

By D. G. AVERY and J. B. GUNN
Radar Research Establishment, Great Malvern, Worcs.

MS. received 13*th May* 1955 *and in final form* 5*th July* 1955

Abstract. A discussion is given of the modifications to the theory of the 'travelling light spot' method for the measurement of minority carrier lifetime τ, which become necessary when the light spot is intensity modulated at an angular frequency ω. It is shown that if $\omega\tau \gg 1$, the 'lifetime' measured is approximately $2/\omega$ and is independent of τ. It is also shown that, if the phase of the collector signal is measured as well as its amplitude, the diffusion constant and hence the drift mobility of minority carriers can be measured directly.

§ 1. Introduction

The use of the 'travelling light spot' method for the measurement of minority carrier lifetimes in semiconductors is now well established as a routine technique (Goucher 1951). In this method, a spot of light is traversed over the surface of the specimen, and the density of optically excited minority carriers is measured as a function of the radial distance of the light spot. The theory of this effect has been derived for static conditions (e.g. Valdes 1952), and shows that the minority carrier diffusion length can be obtained from the graph of minority carrier density against distance, and that, if the diffusion constant is known, the lifetime can be calculated.

In practice, the minority carrier density is measured by observing the change in potential of a collector contact. In order to simplify the problem of amplifying this small change, it is the universal practice to modulate the light spot, so permitting the use of an a.c. amplifier. This modification introduces a previously unrecognized limitation on the theory as given, and also allows additional information to be obtained from the experiment by observing the phase, as well as the amplitude, of the collector signal.

§ 2. Theory

The theory given below is derived on the assumption that the dimensions of the light spot are much less than its separation from the collector, that surface recombination effects are negligible, and that no light of wavelength near the absorption edge of the semiconductor is scattered from the spot (Arthur *et al.* 1955).

The derivation is given for highly extrinsic n-type material, in which the ambipolar diffusion constant D is equal to D_p, the value for holes. It may obviously be extended to material of any type by the substitution of the appropriate value of D.

If p is the hole density at radius r from the light spot (of radius r_0), and τ is the (constant) hole lifetime, then in the present case p is that solution of the equation

$$\frac{D_p}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p}{\partial r}\right)=\frac{\partial p}{\partial t}+\frac{p-p_\infty}{\tau}$$

which satisfies the boundary conditions

$$p-p_\infty=P_0\{1+\exp(i\omega t)\}\text{ at }r=r_0$$

and

$$p\to p_\infty\quad\text{as}\quad r\to\infty.$$

The solution may be written in the form $p-p_\infty=P(r)\exp i\omega t+Q(r)$ and is

$$\frac{P}{P_0}=\frac{r_0}{r}\exp\left\{\left(\frac{1+i\omega\tau}{D_p\tau}\right)^{1/2}(r_0-r)\right\}$$

$$\frac{Q}{Q_0}=\frac{Q}{P_0}=\frac{r_0}{r}\exp\left\{\frac{r_0-r}{(D\ \tau)^{1/2}}\right\}.$$

($2Q$ is, of course, the static solution ($\omega=0$).)

On expressing P/P_0 in terms of the real variables R, θ, as $P/P_0=R\exp(i\theta)$, we have

$$\theta=\frac{r_0-r}{(2D_p\tau)^{1/2}}\{(1+\omega^2\tau^2)^{1/2}-1\}^{1/2} \qquad \ldots\ldots(1)$$

and

$$\ln\frac{Rr}{r_0}=\frac{r_0-r}{(2D_p\tau)^{1/2}}\{(1+\omega^2\tau^2)^{1/2}+1\}^{1/2}. \qquad \ldots\ldots(2)$$

It should be noted that *only when $\omega^2\tau^2$ is much less than* 1 is R approximately equal to Q/Q_0. Since it is always R which is measured in the travelling light spot experiment, very considerable inaccuracy may be introduced by interpreting the results as measurements of Q when this condition is not fulfilled. In the limiting case of $\omega^2\tau^2 \gg 1$, the failure to distinguish between R and Q will lead to a measured 'lifetime' τ', which is given by $\tau' \sim 2/\omega$, and which bears no relation to the true value τ.

The additional information which is obtained when θ is measured as well as R may be utilized by differentiating equations (1) and (2) with respect to r, and taking the product and ratio of the resulting equations. The results are:

$$\frac{d\theta}{dr}\frac{d\{\ln(Rr/r_0)\}}{dr}=\frac{\omega}{2D_p} \qquad \ldots\ldots(3)$$

and

$$\frac{d\{\ln(Rr/r_0)\}}{d\theta}=\frac{(1+\omega^2\tau^2)^{1/2}+1}{\omega\tau}. \qquad \ldots\ldots(4)$$

The calculation of τ from equation (4) will be most accurate in the case where $\omega^2\tau^2 \ll 1$, since in this case the expression for τ does not contain the small difference of large quantities. Subject to this restriction, equation (4) may be re-written as

$$\frac{d\{\ln(Rr/r_0)\}}{d\theta}\sim\frac{2}{\omega\tau}. \qquad \ldots\ldots(4a)$$

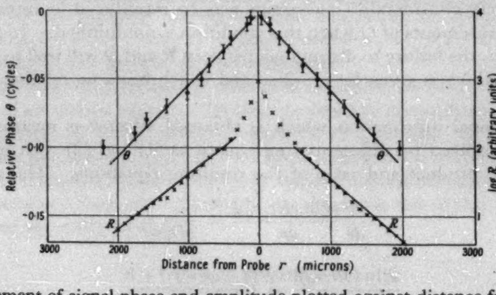
The value of τ so obtained is correct within 10% if $d\{\ln(Rr/r_0)\}/d\theta \geqslant 4$. In practice, the difference between $d\{\ln (Rr/r_0)\}$ and $d(\ln R)$ is usually so small that the latter quantity may be used in equations (3) and (4*a*) without loss of accuracy.

It will be seen from equations (3) and (4) that it is possible to measure both the lifetime *and the diffusion constant* of minority carriers by means of a travelling light spot experiment. This result, together with the Einstein relation, $D=\mu \mathbf{k} T/q$, should permit the drift mobility μ of minority carriers to be measured in those semiconductors in which the classical Haynes–Shockley method is difficult or impossible. Even the collector contact may be dispensed with, since the local minority carrier density can be measured by a non-rectifying probe, which assumes a floating potential dependent on the carrier density. Alternatively, prior knowledge of D may be used as a check on the accuracy of lifetime measurements, since an experiment which is subject to any of the usual causes of inaccuracy (collector non-linearity, scattered light, etc.) will yield the wrong value of D.

§ 3. EXPERIMENT

Equation (3) has been verified by measurements made on a sample of n-type germanium of resistivity about 8 ohm cm. A spot of light about 15 μ in diameter was formed on the freshly etched surface by means of a reflecting optical system. The intensity of the spot was modulated at 1000 c/s by a chopper disc, and its position could be measured to $\pm 1\,\mu$. A hole density signal was obtained from an unbiased, unformed probe of tungsten wire, and was amplified and detected in a homodyne system. The reference signal for the homodyne detector was derived optically from the chopper disc, and could be varied in phase by a simple mechanical arrangement.

The results of a typical experiment are shown in the figure, from which may be derived the values $d\theta/dr = 3{\cdot}45$ radian cm^{-1} and $d(\ln R)/dr = 20{\cdot}0$ cm^{-1}. From equations (3) and (4*a*) the parameters of this specimen are thus found to be $\tau = 54\,\mu$sec and $D_p = 45$ cm^2 sec^{-1}. The drift mobility of holes in material of this resistivity is 1950 cm^{2}v^{-1} sec^{-1} at 290°K (Prince 1953), corresponding to $D = 50$ cm^2 sec^{-1}.



Measurement of signal phase and amplitude plotted against distance from probe.

This measure of agreement is considered satisfactory, in view of the difficulty of measuring the small phase shifts involved. A higher chopping frequency would obviously have been desirable in this case.

ACKNOWLEDGEMENTS

The authors are indebted to their colleagues at Radar Research Establishment for help and encouragement, and in particular to Mr. P. Porteous for performing the measurements, Mr. D. A. H. Brown for the design of the homodyne system, and Dr. A. F. Gibson for stimulating discussions.

REFERENCES

ARTHUR, J. B., BARDSLEY, W., GIBSON, A. F., and HOGARTH, C. A., 1955, *Proc. Phys. Soc.* B, **68,** 121.

GOUCHER, F. S., 1951, *Phys. Rev.*, **81,** 475.

PRINCE, M. B., 1953, *Phys. Rev.*, **92,** 681.

VALDES, L. B., 1952, *Proc. Inst. Radio Engrs*, N.Y., **40,** 1420.

The Influence of Pressure on Metal–Germanium Contacts

By J. LEES and S. WALTON
Research Laboratories of The General Electric Company, Ltd.,
Wembley, Middlesex

MS. received 25th March 1955 *and in amended form 7th July* 1955

Abstract. The dependence of the electrical properties of the germanium immediately below a point contact on the stress existing in this region is investigated. Electrical and optical methods of measuring the area of these contacts gave different results. This anomaly is shown to arise from the variation of the resistance of the germanium with pressure.

---

§ 1. Introduction

It is well known that the theoretical treatment of the d.c. voltage characteristics of metal point contacts on to a semiconductor such as germanium or silicon does not lead to quantitative agreement with experimental results. Various attempts have been made to modify the simple barrier model, the most important of which are the theory of surface states (Bardeen 1947) and the inclusion of minority carrier injection (Banbury 1953).

In this paper it is shown that, in a theoretical treatment of the problem, a further effect may be of importance. This effect is associated with a change in the electrical properties of the semiconductor just beneath the contact point, due to the high mechanical stresses which exist in this region when metal point contacts are applied in the normal way. The phenomenon has been studied using contacts on to germanium of such resistivity and surface treatment that the barrier resistance is negligible in comparison with the spreading resistance of the contact.

The dependence of the electrical properties of the germanium immediately beneath the contact on the stresses existing in this region is deduced from an anomaly which is found to exist in the values of the area of contact determined by two different methods. The principles of these methods are outlined below, and in § 3 it is shown that they are applicable to contacts on germanium. An account is then given of the results obtained when tungsten and various other metals are used as contacting electrodes.

§ 2. Determination of Contact Areas

The area A of a flat circular metal contact on a semi-infinite material is related to the spreading resistance R_s associated with the contact by the relation

$$R_s^2 = \pi\rho^2/16A \qquad \ldots\ldots(1)$$

where ρ is the resistivity of the material beneath the contact. As described elsewhere (Lees and Walton 1955), certain resistivities of p-type germanium may be prepared on which point contacts produce negligible rectifying barriers, and for these contacts R_s, and thus A by equation (1), may be obtained directly from the slope of the d.c. voltage characteristic.

An estimate of the contact area may also be made from the mechanical properties of the contact. When a pyramidal or conical shaped tip of a solid is pressed into a softer material, plastic deformation of the latter occurs. If the included angle of the tip is between 120° and 180°, deformation continues until the mean pressure P_{m} over the contact is given by (Tabor 1951)

$$P_{\mathrm{m}} \simeq 3Y \qquad \ldots\ldots(2)$$

where Y is the elastic limit of the softer material in pure tension. Where, on the other hand, the pyramid or cone is the softer material, the included angle again being between 120° and 180°, we have

$$P_{\mathrm{m}} \simeq 2{\cdot}7Y. \qquad \ldots\ldots(3)$$

In both of these cases, the pressure P_{m} is independent of the load since geometrically similar deformations are produced at any load.

Use of these considerations is made in the Vickers pyramid hardness test in which an indentation is made in the material to be examined by means of a diamond pyramid applied with a given load. The angle between the opposite faces of the pyramid is 136°. The Vickers pyramid hardness number (V.P.N.) of the material is then defined as the load divided by the total surface area of the pyramidal indentation: it is usually expressed in kg mm^{-2}. By simple geometry, the mean pressure over the contact is given by

$$\text{V.P.N.} \simeq 0{\cdot}93 P_{\mathrm{m}} \qquad \ldots\ldots(4)$$

and using equation (2) we have

$$\text{V.P.N.} \simeq 2{\cdot}8Y. \qquad \ldots\ldots(5)$$

Thus for a metal in the shape of a cone of included angle between 120° and 180°, which is deformed against a harder material, the mean pressure over the contact is given by equation (3): equation (5) shows that the V.P.N. of the metal is approximately equal to the mean pressure over the contact.

The value of P_{m} for a contact between germanium and a 150° cone of a softer metal (this includes all the commonly used electrode metals) is thus approximately equal to the V.P.N. of the metal, and we may derive the contact area knowing the load applied to the contact.

§ 3. PRELIMINARY EXPERIMENTS

3.1. *Spreading Resistance of Electroplated Contacts*

In order to check the spreading resistance formula with stress-free contacts, the surface of a 6·6 ohm cm p-type germanium specimen was plated electrolytically with indium. Spots of masking agent were applied to the surface and the surrounding indium dissolved away in dilute nitric acid. The masking agent was removed, leaving a series of indium contacts. The areas of these contacts, measured optically, varied from $1{\cdot}7 \times 10^{-4}$ cm^2 to $3{\cdot}7 \times 10^{-3}$ cm^2. The electrical characteristics of these contacts were obtained in the range ± 1 volt, and were quite linear and symmetrical. Barrier effects were thus negligible in this voltage range. The spreading resistance R_{s} of each contact was obtained from the slope of its electrical characteristic. Figure 1 shows a plot of $1/R_{\mathrm{s}}^2$ against the area A measured optically. Applying equation (1), the slope of the straight line obtained yields a value of 6·5 ohm cm for the resistivity of the germanium. This agrees, within the experimental error, with the resistivity measured by the four-probe technique (Valdes 1954).

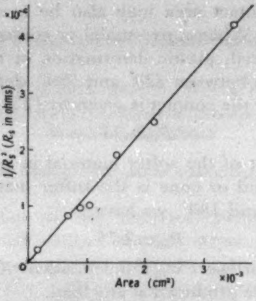


Figure 1. $1/R_s^2$ plotted against optically measured area A for indium plated contacts on p-type germanium.

3.2. *Contact Areas derived from Hardness*

It was not possible to measure the hardness of germanium using the conventional Vickers testing machine since the applied loads (~1 kg) cause extensive featuring around the indentation. However, by using a diamond indenter with loads up to 100 g, good indentations were obtained in the germanium and these corresponded to a V.P.N. of 610 kg mm^{-2}.

According to § 1, the mean pressure over a deformed cone pressed against a harder material is approximately equal to the V.P.N. of the cone. This conclusion was investigated experimentally. The V.P.N's of tungsten, molybdenum, brass and platinum, in the form of 2 mm rods, were measured. One end of each rod was ground to a 150° cone and applied to germanium with loads of 2150, 4150 and 6150 grammes. The metal rods were removed after each loading and the radii of the flattening of the conical tips were measured optically. In table 1 these radii are compared with the values expected from the V.P.N. of each metal: agreement is obtained within experimental error. As a result of elastic recovery when the metal rods are removed from the germanium for measurement, the true radii of contact under load tend to differ from the values given in table 1. Metal specimens were pressed with various loads against a thin sapphire disc and the true radii of contact observed through the sapphire. These radii were not found to be significantly different from the values given in table 1.

Table 1. Observed and Calculated Values (in μ) of Contact Radius

Metal	V.P.N. (kg mm^{-2})	Load 2150 g		Load 4150 g		Load 6150 g	
		Calc.	Obs.	Calc.	Obs.	Calc.	Obs.
Tungsten	500	37	39 ± 4	51	52 ± 4	63	64 ± 4
Molybdenum	300	48	47	66	70	81	82
Brass	172	63	64	88	87	107	102
Platinum	97·5	84	77	116	114	142	143

§ 4. TUNGSTEN–GERMANIUM CONTACTS

Single crystal p-type germanium of resistivity 6·5 ohm cm was used in this experiment. The surface was prepared by etching in CP4 (Haynes and Shockley 1951). The tungsten specimens were in the form of (i) a 0·2 mm diameter wire cut electrolytically to a 150° cone, and (ii) a 2 mm rod, ground to a 150° cone.

The wire was applied to the germanium with loads of from 3 g to 65 g and the rod with loads of from 150 g to 6150 g. The d.c. voltage characteristics were obtained at each load and were approximately linear and symmetrical in the range ± 0·8 volt (figure 2). Values of the spreading resistance were used in equation (1) to calculate the apparent contact area A_s as a function of load. The results are shown in figure 3. A_s is seen to be a substantially linear function of the load over the entire loading range of from 3 g to 6150 g. The slope of this function, which is 64 kg mm^{-2}, gives the apparent mean pressure over the contact. The mechanical considerations dealt with in §§2 and 3 indicate that the true mean pressure over the contact is approximately equal to the V.P.N. of the tungsten. This was measured and found to be 500 kg mm^{-2}. The contact areas A_s derived from the spreading resistance measurements are thus a factor of 7·8 times larger than the true contact areas.

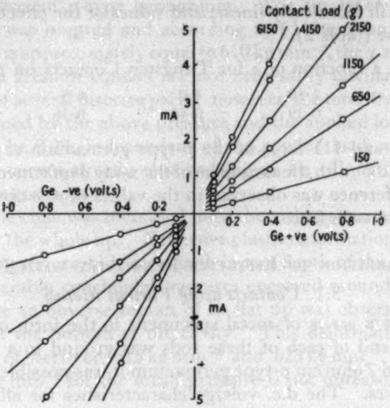


Figure 2. Characteristics of a tungsten contact on p-type germanium.

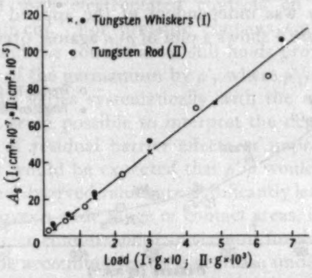


Figure 3. Load plotted against apparent area for tungsten contacts.

Considering equation (1), we see that this anomaly may be described by using a value of the resistivity which is smaller than the true value ρ. This smaller value we call the effective resistivity and denote by the symbol ρ'. For the tungsten contacts above, the value of ρ'/ρ is 0·36.

With the larger area contacts such as are obtained using the 2 mm rod loaded up to 6150 g, barrier effects which depend on $(\text{area})^{-1}$ tend to become small in comparison with the spreading resistance which depends on $(\text{area})^{-1/2}$. The d.c. voltage characteristics of such contacts on a range of resistivities of p-type germanium were in fact nearly linear, and values of the effective resistivity were found in each case (table 2).

Table 2. ρ'/ρ as a function of ρ for Tungsten Contacts on p-type Germanium

ρ(ohm cm)	1·10	4·0	6·5	26
ρ'/ρ	0·39	0·40	0·36	0·38

The 100, 110 and 111 faces of the p-type germanium crystal were exposed with an accuracy to 2° with the assistance of the x-ray department of the laboratory. No significant difference was observed in the values of ρ'/ρ obtained for tungsten contacts applied to each face.

§ 5. VARIATION OF EFFECTIVE RESISTIVITY WITH PRESSURE

5.1. *Contacts using Various Metals*

The V.P.N's of a series of metal specimens in the form of 2 mm rods were measured. One end of each of these rods was ground to a 150° cone and by applying them to 6·2 ohm cm p-type germanium it was possible to achieve a range of contact pressures. The d.c. voltage characteristics for all of these contacts were linear. The values of spreading resistance, together with the true area of contact derived from the hardness of each metal, were used in equation (1) to obtain the effective resistivity at four different loads for each metal. In each case the effective resistivity was independent of the applied load and therefore of the contact area. Figure 4 shows a plot of ρ'/ρ against pressure at these contacts.

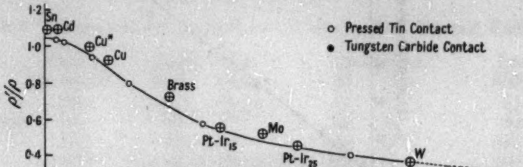


Figure 4. Effective resistivity plotted against pressure.

5.2. *Constant Area Contacts*

The end of a tungsten carbide rod was ground to a truncated cone, the area of the flat tip being 10^{-4} cm^2. The cone was punched through a thin tin foil and this procedure resulted in a thin layer of tin covering the tip. When this tip was applied to 6·2 ohm cm p-type germanium, the layer of tin ensured that the whole area of the tip was in contact with the germanium for each of the applied loads of from 150 g to 4150 g. Tungsten carbide is considerably harder than germanium, thus in this case increase in load results in an increase in pressure at the contact. The electrical characteristics at each load were linear, and from the values of spreading resistance obtained the effective resistivity was derived as a function of contact pressure (figure 4).

5.3. *Tungsten Carbide Contacts*

A truncated tungsten carbide cone was applied with loads of from 1150 g to 6150 g to 6·2 ohm cm p-type germanium. The tin layer described in the preceding section was omitted and according to § 1 the mean pressure over the contact in this case is approximately equal to 610 kg mm^{-2}, the V.P.N. of germanium. The shapes of the areas of contact are not necessarily circular or elliptical and may even consist of several discrete parts; however, the total area of any particular contact is determined by the above pressure and the applied load. The value of ρ'/ρ was found to be 0·34, independent of the applied load, and lies on the extrapolation of the curves already obtained (figure 4). Equation (1) is thus still applicable to these contacts.

The load on the cone was increased until the yield pressure of the germanium was exceeded over the whole tip. Extensive plastic deformation of the germanium took place, the tungsten carbide cone sank in to a depth of the order of one millimetre and considerable conchoidal fractures occurred around the contact. On removing the cone an obverse replica of its flat tip was observed impressed into the germanium at the bottom of the cavity. This deformation took place at a contact pressure of between 615 kg mm^{-2} and 625 kg mm^{-2}, showing that our estimate of 610 kg mm^{-2} for the mean pressure is not unreasonable.

§ 6. DISCUSSION

The results presented in § 3 indicate that equation (1) holds for spreading resistances associated with electroplated contacts on p-type germanium of resistivity about 6 ohm cm. For pressure contacts equation (1), relating the spreading resistance and the contact area, still holds provided that we replace ρ the normal resistivity of the germanium by ρ', where ρ'/ρ is independent of load for a given metal, and varies systematically with the mean pressure over the contact. It does not seem possible to interpret the departure of the ratio ρ'/ρ from unity in terms of residual barrier effects or imperfect electrical contact, since in these cases it would be expected that ρ'/ρ would be greater than unity, whereas, in general, the observed values are significantly less than unity. Further, since ρ'/ρ is constant over a wide range of contact areas, it appears most unlikely that an additional surface conductance can account for the observed results.

The results may be accounted for by a decrease under pressure of the actual resistivity of the germanium under the contact, that is in the region which determines the value of the spreading resistance. The effective resistivity ρ' is then a measure of the resistivity change which has taken place.

For tungsten contacts it appears that ρ'/ρ is independent of ρ over a range of p-type resistivities. In figure 4 the two methods of obtaining ρ'/ρ as a function of pressure are in good agreement; a smooth variation of ρ'/ρ with pressure is found from a value of approximately 1·0 at low pressures, to the value of 0·34 at 610 kg mm^{-2}, the yield pressure of germanium. For pressures greater than 100 kg mm^{-2}, the curve may be represented by the empirical relation (pressure)$\rho'^2 = 3 \times 10^8$ (g ohm^2). It is interesting to note that this square law relation implies that for the harder metal contacts, that is contacts for which the mean pressure exceeds 100 kg mm^{-2}, the spreading resistance at a given load is independent of the metal used. The points in figure 4 corresponding to low pressures tend to give values of ρ'/ρ which are slightly greater than unity. This effect is probably due to imperfect contact at low pressures and to small variations ($\sim$5%) in resistivity ρ over the germanium surface.

The effect of pressure on the resistivity of germanium has been investigated by several workers. Bridgman (1950) and Paul and Brooks (1954) have subjected germanium to hydrostatic pressures up to 300 kg mm^{-2}. The resistivity of p-type material was found to decrease with pressure but the magnitude of the effect was much smaller than that which we have observed. True hydrostatic pressure is not, however, realized beneath our contacts. Smith (1954), working at lower pressures ($\sim$1 kg mm^{-2}), has demonstrated the existence of a large shear coefficient of resistivity which, when extrapolated to the pressures used in our experiments, would result in a change in resistivity of the same order and in the same direction as we have observed. This effect is, however, dependent on crystal orientation. Our experiments have not indicated a dependence of ρ'/ρ on crystal direction (§ 3). Further investigations into the stress distribution under a point contact on to germanium are required before an adequate comparison may be made between our results and those of Smith. Work is being continued to extend measurements of effective resistivity to metal contacts on n-type germanium.

REFERENCES

BANBURY, P. C., 1953, *Proc. Phys. Soc.* B, **66,** 833.
BARDEEN, J., 1947, *Phys. Rev.*, **71,** 717.
BRIDGMAN, P. W., 1950, *Proc. Amer. Acad. Arts Sci.*, **79,** 122.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81,** 835.
LEES, J., and WALTON, S., 1955, *Proc. Phys. Soc.* B, **68** (in the press).
PAUL, W., and BROOKS, H., 1954, *Phys. Rev.*, **94,** 1128.
SMITH, C. S., 1954, *Phys. Rev.*, **94,** 42.
TABOR, D., 1951, *The Hardness of Metals* (Oxford: University Press), ch. 9.
VALDES, L. B., 1954, *Proc. Inst. Radio Engrs, N.Y.*, **42,** 420.

Absolute Scintillation Efficiency of Anthracene Crystals

By G. T. WRIGHT†

Physics Department, Rhodes University, Grahamstown, South Africa

MS. received 10*th May* 1955 *and in amended form* 21*st June* 1955

Abstract. The absolute scintillation efficiency of anthracene crystals excited with fast electrons is measured. The experimental method used avoids the errors existing in previous measurements which are due to the anisotropic refraction, spatial distribution and re-absorption of fluorescence in these crystals. At a temperature of 290°K the intrinsic scintillation efficiency (in the absence of re-absorption of fluorescence) is found to be $0{\cdot}062 \pm 0{\cdot}004$; this is reduced to $0{\cdot}050 \pm 0{\cdot}005$ for the crystal by self-absorption of fluorescence.

Reasons for this low efficiency are discussed. A detailed description is given of excitation conditions produced in organic crystals by ionizing particles with a discussion of possible quenching mechanisms in the excitation column. For electrons these are considered to be due to internal molecular dissipation of electronic excitation energy, to mutual interaction between excited molecules and to the strong electric fields surrounding ions. For the heavy particles these processes are more pronounced and further quenching occurs due to thermal fields and to damaged molecules. The role of migration of excitation energy is also considered.

This description of the scintillation process satisfactorily explains the low efficiencies and the decay shapes of fluorescence observed experimentally.

§ 1. Introduction

In a large number of organic crystals the passage of an ionizing particle is accompanied by fluorescence emission. This phenomenon has acquired considerable practical importance as the basis of the photomultiplier scintillation counter although its underlying physical mechanisms remain little understood. Of particular interest due to its fundamental position in leading to an understanding of the processes involved is the efficiency with which the energy absorbed from the particle is converted into energy of fluorescence. For the case of anthracene excited with fast electrons this has been estimated by several experimenters (Hopkins 1951, Harrison 1952, Furst, Kallmann and Kramer 1953, Birks and Szendrei 1953) who have reported values between 0·01 and 0·1. Notwithstanding the differences in these results this is very low compared with the photofluorescence efficiency which is 0·80 for excitation with light of the same wavelength as the mean fluorescence wavelength (Wright 1955 a). Any theory of the scintillation process must provide a satisfactory explanation for its low efficiency and it is consequently most desirable to obtain a reliable measurement of this.

The major errors in a direct measurement occur in estimating the fraction of the fluorescence light received by the detecting instrument. The fluorescence intensity is measured directly only within a certain solid angle external to the crystal; this has then to be related, by calculation, to the total fluorescence intensity within the crystal. Since this is very anisotropic as regards both refraction and spatial distribution of fluorescence in addition to showing considerable re-absorption of fluorescence, many assumptions and approximations are necessary for this purpose and introduce considerable uncertainty into the final result. It is mainly for this reason that a reliable measurement of the scintillation efficiency has not yet been made for any substance.

---

† Now at Mullard Research Laboratories, Salfords, Redhill, Surrey.

The present communication describes a measurement of the absolute scintillation efficiency of anthracene crystals excited with fast electrons. This substance was chosen because of its importance in the field of scintillation counting. Further, the absolute photofluorescence quantum efficiency of this crystal has recently been determined. This enables a direct measurement to be made of the efficiency with which fluorescence photons are collected by the detecting instrument and thus removes the major source of error existing in previous measurements.

§ 2. EXPERIMENTAL

The experimental method consisted of using a photomultiplier, calibrated with a known radiant energy flux, to measure the fluorescence energy received from an anthracene crystal excited with a known energy flux of electrons. The efficiency of collection of fluorescence photons could be measured and consequently the absolute fluorescence energy emitted within the crystal could be found. Since the incident electron energy flux and the emitted fluorescence energy flux were known the absolute scintillation efficiency was obtained.

A more detailed description of the experimental method may be conveniently divided into three sections dealing respectively with calibration of the photon source, calibration of the electron source and measurement of the efficiency of collection of fluorescence photons.

2.1. *The Photon Source*

A mercury lamp with an envelope of Wood's glass was used as the primary source of ultra-violet radiation. The radiations from this lamp were further filtered through Chance OB8 and OX1 glass to isolate the 3650 Å group of lines and emerged through an aperture of known area about half a centimetre in diameter. The radiant intensity in the plane of this aperture was measured with a flag thermocouple by comparison with the total radiant intensity emitted from a standard tungsten filament lamp supplied and calibrated by the National Physical Laboratory, England. Slow drifts in the radiation output of the mercury lamp were indicated by a photocell and were compensated by manual control of the lamp voltage. After calibration the energy flux of 3650 Å radiation emerging from the lamp aperture was reduced with three neutral filters. This combination had been previously calibrated for this wavelength and reduced the intensity by a known factor of the order of 10^6. This calibration of the photon source was repeated at the conclusion of the scintillation efficiency measurements.

2.2. *The Electron Source*

The 624 kev conversion electrons from ^{137}Cs were used for this purpose. Since this source emits many more gamma rays than electrons, however, a polonium alpha particle source was used to excite the fluorescence of the crystal in the first instance.

Fluorescence intensities were measured in terms of the output current of the calibrated photomultiplier. Following each measurement of the fluorescence intensity excited by the alpha particles pulse height distribution curves were measured for both the ^{137}Cs and ^{210}Po sources and the pulse height ratio for 624 kev electrons and polonium alpha particles determined. Knowing this ratio and knowing the fluorescence intensity excited by a given numerical flux of alpha particles, the fluorescence intensity that would be excited by the same numerical flux of 624 kev electrons could be determined. This method has the merit that it is not necessary to know the energy of the alpha particles incident on the crystal. This was estimated as 3·80 MEV, however, in order to compare the ratio measured with that reported by other authors. Each of the crystals used in the present measurement gave an electron to alpha particle pulse height ratio of 2·6. This agrees satisfactorily with a ratio of 2·5 obtained from the measurements of Taylor *et al.* (1951) and a ratio of 2·8 derived from the measurements of Hopkins (1951).

2.3. *Efficiency of Collection of Fluorescence*

This measurement has previously been described in detail (Wright 1955 b) consequently only a brief description will be given here.

When the crystal is illuminated with light in its absorption region this is converted into fluorescence with a quantum efficiency less than unity and only a part of the fluorescence is collected by the photomultiplier. Thus the multiplier output current is less than when the exciting light is received directly by the photocathode. The ratio of the multiplier output currents, when corrected for the different spectral sensitivity of the photocathode to the exciting light and the fluorescence light, evidently gives the product of the fluorescence quantum efficiency and the fluorescence collection efficiency. The quantum efficiency of anthracene is known, hence the collection efficiency may be obtained. For the particular experimental conditions used the mean collection efficiency was 0·47.

§ 3. EXPERIMENTAL PROCEDURE

The anthracene crystals used for this measurement were cubes of side 1 cm supplied for scintillation counting by the Larco Nuclear Instrument Company. A thin Perspex cap was shaped to fit the cathode window of the RCA type 5819 photomultiplier tube used to detect the fluorescence and the crystal placed on top of this. A screen cut from black paper was placed over the photocathode to ensure that this received only fluorescence transmitted through the base of the crystal. Optical contact between surfaces was improved by the introduction of thin films of glycerine. Before mounting the crystal its top face was polished gently on soft tissue paper. The portion of the crystal face exposed to the incident radiations was limited to a circular area of 5 mm diameter by a thin foil diaphragm.

The efficiency of collection of fluorescence was first measured, as described in the previous section, followed by measurement of the multiplier current produced by excitation of the crystal with alpha particles. Pulse height spectra were then taken for the alpha particle and electron sources. To ensure linear response of the multiplier the instantaneous pulse current was not allowed to exceed 1 mA. This sequence of measurements was made for excitation of each of three adjacent mutually perpendicular faces on each of two crystals which possessed very few flaws. Between changing crystal faces the multiplier current was measured with the photocathode receiving directly the known energy flux of 3650 Å radiation. This measurement, combined with the known spectral sensitivity curve of the photocathode which had been previously measured with a Hilger type F.T.16 linear vacuum thermopile (Wright 1955 b), provided an absolute calibration of the photomultiplier. Using this calibration curve the fluorescence energy received by the photocathode was determined. Since the fluorescence collection efficiency was known the total fluorescence energy excited in the crystal with the alpha particles, and hence the electrons, could be found and the absolute scintillation efficiency evaluated.

§ 4. EXPERIMENTAL RESULTS

Each crystal gave the same result for the absolute scintillation efficiency. For excitation with fast electrons the scintillation efficiency of the molecular fluorescence at a temperature of 290°K was found to be $0{\cdot}062 \pm 0{\cdot}004$. This is the efficiency which the crystal would have if anthracene did not re-absorb part of its own fluorescence. Re-absorption of fluorescence in the crystal increases the mean fluorescence wavelength from 4200 Å for the molecular fluorescence to 4450 Å for the crystal fluorescence and reduces the quantum efficiency from 0·94 to 0·80 (Wright 1955 a). Consequently the scintillation efficiency of the crystal fluorescence is less than that of the molecular fluorescence and is $0{\cdot}050 \pm 0{\cdot}005$. The uncertainty quoted in these results is due mainly to experimental error in calibrating the photomultiplier and to uncertainty in the photofluorescence quantum efficiency.

An alternative way of stating these results is that for excitation of the molecular fluorescence the energy expenditure by the absorbed particle is 47 ev for each photon emitted; for the crystal fluorescence the energy expenditure is 55 ev for each photon emitted.

These measurements verify that the scintillation efficiency of anthracene is very much less than the photofluorescence efficiency. Since all organic crystals have a similar response this low efficiency is a general characteristic of the scintillation process. Reasons for this low efficiency are discussed in the next section which presents a description of the scintillation process based on a broad division into three parts. The first of these describes excitation conditions produced by ionizing particles; this is followed by a tentative description of quenching and migration of excitation energy and the last part deals with the final process of fluorescence emission.

§ 5. THE SCINTILLATION PROCESS

5.1. *Excitation Conditions produced by Ionizing Particles*

The excitation conditions produced by ionizing particles differ considerably from those prevailing during photo-excitation and may be expected to modify the nature of the fluorescence emission; a detailed consideration of excitation conditions is thus of interest.

The incident particle loses its energy by ionization and excitation of the molecules along or near its path. The ionization and excitation densities produced are greater the slower the particle and hence increase along the track. In the primary excitation column so formed molecules merely raised into excited electronic levels outnumber ionized molecules two or three fold. Towards the end of the track energy loss by collision with the crystal lattice, with formation of various lattice defects, becomes appreciable (Seitz 1949). In ionization the energy taken up by a molecule will not exceed the ionization energy of the electron ejected except in the rare case of a direct collision with a nucleus, but the ejected electron can have all energies up to a maximum corresponding to a head-on collision with the incident particle. For a 5 MEV alpha particle this corresponds to a secondary electron energy of 3 kev; for a beta particle the maximum energy of ejection is equal to the energy of the beta particle. Such collisions are rare, however, and electrons ejected at low speeds greatly preponderate. The faster of the secondary electrons produce further ionization and most produce further excitation. It has been established that for 5 MEV alpha particles approximately 60% of the total ionization is produced by the secondary electrons and that the mean energy dissipated for each ion pair formed is about 30 ev. This latter figure varies slightly depending on the nature and physical state of the absorbing system but does not appear to depend on the nature or the velocity of the ionizing particle.

It is not possible to make more than a qualitative estimate of the diameter of the primary excitation column. Direct excitation of molecules by the incident particle will occur at some distance from the actual track. Approximate calculations for a 5 MEV alpha particle indicate a distance of a few molecular diameters for excitation of the molecule into its first electronic level (Mott 1952). As the velocity of the particle decreases this effective collision distance decreases. Diffusion of the secondary electrons produces ionized and excited molecules at some distance from the track of the incident particle but it is difficult to estimate the penetration distances of such low energy electrons. For the great majority of secondary electrons ejection energies are less than 100 ev and, since a 1 kev electron has a range of only about 10^{-6} cm, they probably travel only a few molecular diameters.

Some of the secondary electrons produced will be ejected from low-lying levels of the atoms and will thus have fairly low energies although the transfer of energy from the particle may be large. Subsequent completion of the atomic shell by capture of an electron will result in emission of one or more X-ray quanta. These will usually be captured and give rise to secondary electrons some distance from the path of the incident particle. This process will be rare, however.

The process of particle absorption takes place in a time negligible compared with that required for ion diffusion and recombination hence immediately after passage of the incident particle its track will consist of a high concentration of positive ions located in a narrow core and a lower concentration of negative ions spread throughout a larger volume. The steep potential gradients will result in a gradual radial diffusion of charge leading eventually to charge neutralization. Because of the finite time which elapses before charge neutralization is complete the various positive and negative ions formed may suffer chemical changes leading to the production of new chemical entities in the crystal lattice having different fluorescence characteristics. In addition, molecules raised into high excited levels may dissociate when their electronic energy is transformed into internal vibrational energy. Such new substances will remain in the lattice and will affect luminescence produced by this and subsequent particles.

In solids, recombination may occur by the Franck–Rabinovitch cage effect of fragments formed from the same molecule by dissociation. It is probable that the number of such temporarily damaged molecules will greatly exceed the number of permanently damaged molecules and exist for an appreciable time.

5.2. *Quenching and Migration of Electronic Excitation Energy*

It has been shown previously (Wright 1953 a) that quenching of fluorescence during the scintillation process can be satisfactorily described in terms of monomolecular and bimolecular processes. The description given above of excitation conditions in the primary excitation column suggests several possible mechanisms for these.

Although final fluorescence occurs from the first excited level of the molecule most of the molecules in the excitation column acquire considerably more energy than this from the passing particle. This excess electronic energy is rapidly dissipated as thermal energy within the molecule and is then conducted away through the crystal lattice. Much of the energy absorbed from the particle is dissipated almost immediately in this way and produces much higher temperatures in the vicinity of the excitation column than in the bulk of the crystal. For alpha particles it is estimated that the initial temperature rise is several hundred degrees. This is exaggerated since no account is taken of excitation of internal molecular vibrations but it does indicate the magnitude of the thermal gradients established. This thermal energy will have a deleterious effect upon the fluorescence of the excited molecules since the fluorescence efficiency of these decreases as the temperature increases. For fast electrons the thermal gradients produced are much smaller and may be neglected but they will account to some extent for the much lower scintillation efficiencies observed for the heavy particles. In this connection it is of interest to estimate the time for which these thermal gradients persist. For this purpose we simplify the crystal to an isotropic simple solid in which the phonons of the lattice vibrations are transferred between adjacent molecules in a random fashion independent of the existing temperature gradient; the mean transfer frequency is considered equal to the lattice frequency. The equation describing the diffusion of phonons through the lattice predicts that the thermal conductivity of the solid is $\nu l^2 c_v \rho/6$. In this expression ν is the lattice frequency (which may be calculated from compressibility data), l is the mean phonon path length, which has been taken as the mean molecular diameter and ρ is the density of the solid. For several organic crystals for which experimental data were available the thermal conductivity was calculated and agreed within a factor of two with the measured value. We may therefore have confidence in the use of this simple model to investigate the diffusion of thermal energy on a molecular scale. In this way it is found that the initial thermal energy in the excitation column is dispersed in a time of the order 10^{-11} second; this is sufficiently long for considerable quenching of excited molecules to occur.

A further mechanism for quenching can exist in the intense electric fields present in the excitation column. For electrons these are of the order of MV cm^{-1} and are higher for the heavy particles which produce greater ionization densities. Perturbation of the electronic configuration of an excited molecule invariably decreases its fluorescence efficiency and it is quite feasible that the strong electric fields surrounding ions will adversely affect the fluorescence of excited molecules in the vicinity. Mutual interaction between excited molecules may also lead to non-radiative dissipation of electronic excitation energy (Black 1953). We may estimate the scintillation efficiency of anthracene excited with fast electrons by considering that most of the initially excited molecules are quenched in these ways. This gives an efficiency of 0·09 which is rather higher than the measured value. However, it is very probable that the energy expenditure required for the formation of an ion pair is greater in the solid than in the gas; this will reduce the estimated efficiency. Further it is by no means certain that the fluorescence efficiency of those excited molecules formed by ion recombination is as high as that observed during photofluorescence. In addition, the fluorescence of those excited molecules formed by initial ion recombination will be reduced to some extent by the electric fields due to the remaining ions. Both of these effects will further reduce the estimated efficiency. In view of the necessarily qualitative nature of this discussion the estimated efficiency agrees satisfactorily with the measured efficiency.

A quenching mechanism not yet considered is due to temporarily and permanently damaged molecules. This has been considered previously (Birks 1953, Wright 1953 b, 1954 a) and is negligible for electrons compared with the heavy particles. For the latter they form an additional source of quenching and will account partially for the very much lower scintillation efficiencies of these.

In all of the above quenching mechanisms migration of the excitation energy is of great importance. Due to the weak lattice binding forces the molecules in an organic crystal are relatively widely spaced and mutual interaction between the molecular electronic orbitals is small. Excitation of a molecule produces initially a compressed molecule which, through coupling of its electronic and atomic configurations, expands until the potential minimum of the excited state is reached. Experiment indicates that the expanded electronic orbital can interact with those of adjacent molecules and result in a transference of excitation energy by a quantum mechanical process (Bowen, Mikiewicz and Smith 1949, Wright 1953 b, 1955 b). The excitation energy thus diffuses through the crystal lattice by repeated transfer between adjacent molecules until emitted radiatively or quenched. This migration process is of considerable importance in the scintillation process since it is in this way that the excitation energy is brought into contact with the various quenching centres such as ions or damaged molecules. It is of particular importance for electrons for which the ionization and excitation densities produced are relatively small.

During the scintillation process excitation energy escapes from the vicinity of the excitation column by fluorescence emission. Part of this fluorescence is re-absorbed in the crystal and results in energy transfer over considerable distances. This is illustrated well by the experiments of Ageno and Cortellessa (1952) who have observed fluorescence in *trans*-stilbene crystals occurring several millimetres away from the point of absorption of the particles. These experiments are not considered evidence for the existence of 'primary' photons as stated by Birks (1953). This re-absorption of fluorescence results in a further slight reduction in the scintillation efficiency of the crystal.

For short range surface incident particles surface escape of fluorescence produces a decrease in scintillation efficiency over and above that due to quenching processes in the excitation column (Wright 1955 c, d). This can modify the shape of the observed scintillation response curves at low particle energies.

This discussion of possible quenching mechanisms refers directly to organic crystals but the conditions existing in liquid solutions are similar. Over distances of a few molecular diameters, as are involved in the primary scintillation processes, the liquid possesses a high degree of crystalline order. In addition, the scintillation process takes place in such a short time that molecular diffusion destroying this short range order will not occur appreciably. The main difference from the case of crystals is in the presence of two types of molecules, the solvent molecules which are responsible for the initial absorption and the solute molecules which are responsible for the final fluorescence (Kallmann and Furst 1950, 1951); transfer of energy from the solvent molecules to the solute molecules will occur during the process of migration of excitation energy. The rather lower scintillation efficiencies of these solutions are due to the reduced fluorescence efficiencies of organic molecules in liquid solution and to increased internal quenching of the excitation energy by the solvent molecule before it is transferred to the solute molecule.

An alternative explanation for the low efficiency of the scintillation process has previously been offered by Birks (1953) in terms of a cascade process of photon absorption and emission which is initiated by emission of high energy primary fluorescence photons from ionized molecules. With certain other assumptions this process can qualitatively explain a wide range of scintillation phenomena but no experimental evidence exists for this primary photon cascade process.

5.3. *Emission of Fluorescence*

In this description of the scintillation process final fluorescence occurs from the first electronic excited level of the molecule giving the same fluorescence spectrum as for photoexcitation; this is found experimentally and was implied during the experimental measurement of the scintillation efficiency.

Although most of the initially excited molecules forming the primary excitation column are quenched as described above a small proportion of these will fluoresce. Because the quenching processes greatly predominate, however, the decay time of this emission will be very short and it will contain only a small part of the total eventual fluorescence. Part of this initial emission will be due to Cerenkov radiation; for particles of moderate energies, however, this accounts for only a small part (of the order of a few per cent) of the total radiation emission. The slower process of ion diffusion and recombination followed by fluorescence of most of the excited molecules so produced results in a longer decay component which contains most of the fluorescence emitted during the scintillation process. The scintillation pulse will thus consist of a short initial spike of fluorescence followed by a longer decay component which contains most of the emission. For crystals such as anthracene the initial spike is reduced appreciably by self-absorption of fluorescence; for *p*-terphenyl this effect is negligible, however (Wright 1954 b, Birks and Wright 1954). Measurements have been made of the decay rates of scintillation pulses (Wright 1955 e) and it is found that they are of this form thus providing strong support for the description of the scintillation process given above.

§ 6. CONCLUSIONS

The efficiency of the scintillation process in organic crystals is very much smaller than that of photofluorescence. In particular, for anthracene excited with fast electrons the intrinsic efficiency of the scintillation process (i.e. in the absence of self-absorption of fluorescence) is only 0·062. This compares with a photofluorescence efficiency of 0·94 (for excitation with light of the same wavelength as the mean fluorescence wavelength).

For electron excitation this low efficiency is considered due to internal molecular quenching of excess electronic excitation energy and to quenching of fluorescence by mutual interaction between excited molecules and by the strong electric fields surrounding ions. For the heavy particles these quenching processes are more pronounced and additional quenching occurs due to the presence of large thermal gradients and damaged molecules in the excitation column.

This explanation of the scintillation process is supported by measurements of scintillation decay rates. These show that, as predicted, the scintillation pulse consists of a short initial spike of fluorescence followed by a longer decay component which contains most of the fluorescence emission.

ACKNOWLEDGMENT

The author wishes to thank the South African Council for Scientific and Industrial Research for a grant to purchase equipment.

REFERENCES

AGENO, M., and CORTELLESSA, G., 1952, *Nuovo Cim.* **9,** 196.

BIRKS, J. B., 1953, *Scintillation Counters* (London: Pergamon Press Ltd.; New York: McGraw-Hill Book Co. Inc.).

BIRKS, J. B., and SZENDREI, M. E., 1953, *Phys. Rev.*, **91,** 197.

BIRKS, J. B., and WRIGHT, G. T., 1954, *Proc. Phys. Soc.* B, **67,** 657.

BLACK, F. A., 1953, *Phil. Mag.*, **44,** 263.

BOWEN, E. J., MIKIEWICZ, E., and SMITH, F. W., 1949, *Proc. Phys. Soc.* A, **62,** 26.

FURST, M., KALLMANN, H., and KRAMER, B., 1953, *Phys. Rev.*, **89,** 416.

HARRISON, F. B., 1952, *Nucleonics*, **10,** 33.

HOPKINS, J. I., 1951, *Rev. Sci. Instrum.*, **22,** 29.

KALLMANN, H., and FURST, M., 1950, *Phys. Rev.*, **79,** 857; 1951, *Ibid.*, **81,** 853.

MOTT, N. F., 1952, *Wave Mechanics* (Cambridge: University Press).

SEITZ, F., 1949, *Disc. Faraday Soc.*, **5,** 271.

TAYLOR, C. J., JENTSCHKE, W. K., REMLEY, M. E., EBY, F. S., and KRUGER, P. G., 1951, *Phys. Rev.*, **84,** 1034.

WRIGHT, G. T., 1953 a, *Phys. Rev.*, **91,** 1282; 1953 b, *Proc. Phys. Soc.* A, **66,** 777; 1954 a, *Phys. Rev.*, **96,** 569; 1954 b, *S. A. Congress of Physics* (Pretoria, July 1954); 1955 a, *Proc. Phys. Soc.* B, **68,** 241; 1955 b, *Ibid.*, **68,** 701; 1955 c, *Phys. Rev.*, in the press; 1955 d, *Ibid.*, in the press; 1955 e, submitted for publication.

Some Factors Influencing Field Emission and the Fowler–Nordheim Law

By T. J. LEWIS
Electrical Engineering Department, Queen Mary College, London

MS. received 27th May 1955, and in final form 20th June 1955

Abstract. Departures from the normal Fowler–Nordheim law which have been observed at fields up to 7×10^7 v cm^{-1} have previously been ascribed to space-charge field distortion. It is shown that certain other effects due to failure of the image law, non-uniformity of work function and to possible surface irregularities could produce similar effects.

---

§ 1. INTRODUCTION

An extensive investigation of field emission has been made by Dyke and his co-workers in which it has been shown that for fields of approximately 10^7 v cm^{-1} the emitted current was in agreement with the Fowler–Nordheim theory corrected for the image force. At higher fields up to 7×10^7 v cm^{-1} the experimentally observed current was progressively less than the predicted values. It was suggested that the departures from the theory were due to space-charge distortion of the cathode field by the high current densities produced. This suggestion was supported by corresponding changes in the emission patterns which are interpreted as showing a more uniform current distribution (Dyke and Trolan 1953, Barbour, Dolan, Trolan, Martin and Dyke 1953). Similar emission patterns have been obtained by Benjamin and Jenkins (1940).

The present contribution does not disagree with these views but draws attention to certain other second-order effects which will occur at such high fields and which may produce significant departures from a Fowler–Nordheim law.

§ 2. THE POTENTIAL BARRIER AT THE SURFACE

The Fowler–Nordheim law for the emitted current density J in amp cm^{-2} can be written in the form (Sommerfeld and Bethe 1933)

$$J = 1{\cdot}55 \times 10^{-6} \frac{E^2}{\phi} \exp\left[\frac{-6{\cdot}85 \times 10^7 \phi^{3/2} f(y)}{E}\right] \qquad \ldots\ldots(1)$$

in which E is the electric field in v cm^{-1}, ϕ is the work function of the emitter in ev and $f(y)$ is a function dependent on the image law, originally calculated by Nordheim (1928) and recently corrected by Burgess, Kroemer and Houston (1953). Equation (1) has been established on the assumptions that it is legitimate to use the image law and that the field E is uniform at least over distances of the order of 10 Å from the surface.

Detailed considerations of the nature of the surface forces such as those made by Slater (1934), Bardeen (1936), and Sachs and Dexter (1950) indicate that it is unlikely that the image law holds within about 3 Å of the surface. Such failure of the image law had also been realized at an early date by Langmuir (1916) and by others. Since the surface barrier through which the electrons must tunnel is reduced as E is increased, equation (1) will become inaccurate when the major portion of the reduced barrier is produced by forces which are not of image type. If an image law is assumed, the peak of the barrier will occur at approximately 5 Å from the surface for $E=1\times10^7\,\mathrm{v\,cm^{-1}}$ and at only 3 Å when $E=5\times10^7\,\mathrm{v\,cm^{-1}}$. Thus in the experiments referred to above the fields were such that the resultant surface barrier would be unlike that expected from an image law. Figure 1 illustrates this and shows that the transparency of the true barrier would be less than that used in obtaining equation (1). As a result the usual linear Fowler–Nordheim plot would be modified over the range AC as in figure 2. Such behaviour has in fact been found by Dyke and Trolan (1953) for clean tungsten emitters. While the experimental results have been explained in terms of space charge, it would seem that the effects described here may not be ignored for the

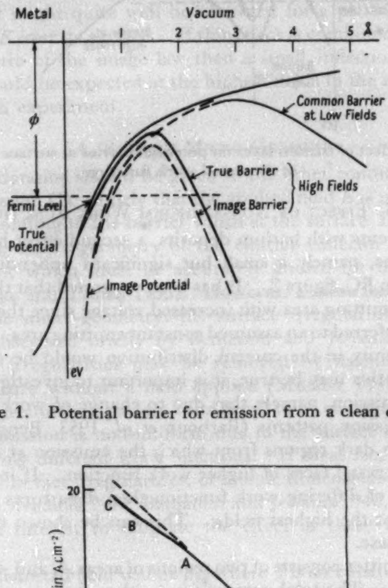


Figure 1. Potential barrier for emission from a clean emitter.

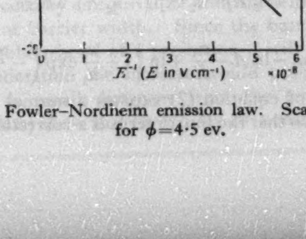


Figure 2. Modified Fowler–Nordheim emission law. Scales only approximate for $\phi=4{\cdot}5$ ev.

field strengths used. Similar behaviour has also been found for a series of tungsten emitters on which barium has been deposited to give a range of effective work functions ϕ between 4·5 ev and 2·44 ev, the effect being more marked and occurring at lower values of E as ϕ is reduced. The present arguments are even more pertinent in such cases since a deposited barium layer in reducing the work function will presumably change the barrier potential considerably over distances corresponding to the layer thickness as in figure 3. For such a barrier a marked departure from equation (1) would be expected as the field is increased, such departure increasing as ϕ is reduced.

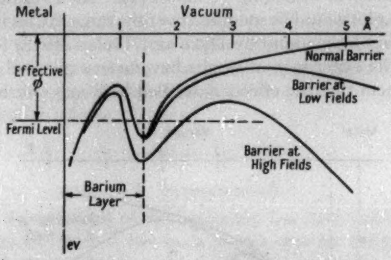


Figure 3. Effect of barium layer on potential barrier at surface showing effect of decreased work function.

§ 3. EFFECT OF NON-UNIFORM WORK FUNCTION

In the experiments with barium deposits, a second effect has been noticed at the highest stresses, namely a small but significant upbending of the current curves in the region BC, figure 2. It has been suggested that this may be due to a slight increase in emitting area with increased voltage since the curves of figure 2 must perforce be referred to an assumed constant emitting area. Such an increase due to non-uniformity in the current distribution would be due to the emitter geometry. While this may be true, it is important to investigate another source of non-uniform emission, namely that due to change of work function over the surface. The emission patterns (Barbour *et al.* 1953, Benjamin and Jenkins 1940) show certain dark regions from which the emission at low fields is slight corresponding to crystal faces of higher work function. If indeed the emission occurs from areas of differing work function then departures from equation (1) may be expected at the highest fields. This can be shown by considering the following simple case.

Suppose the emitter consists of two regions of areas A_1 and A_2 of work function ϕ_1 and ϕ_2 ($\phi_1 < \phi_2$). The total current from either region can be written in the form $I = a \exp(-\alpha/E)$ in which $a = 1{\cdot}55 \times 10^{-6} \phi^{-1} E^2 A$ and $\alpha = 6{\cdot}85 \times 10^7 \phi^{3/2} f(y)$. The average current density J for the two sites is then given by

$$JA = a_1 \exp(-\alpha_1/E) + a_2 \exp(-\alpha_2/E) \quad \text{where} \quad A = A_1 + A_2.$$

Thus

$$\ln J = \ln a_1 - \ln A - \frac{\alpha_1}{E} + \ln\left\{1 + \frac{a_2}{a_1}\exp\left(-\frac{\alpha_2 - \alpha_1}{E}\right)\right\} \qquad \ldots\ldots(2)$$

The first three terms of equation (2) express a normal Fowler–Nordheim law similar to equation (1) so that the fourth term is a correction factor modifying this law. If $\phi_1=\phi_2$, equation (2) reverts to equation (1) with a correct value for the total emitting area. The magnitude of the correction obviously increases with E and is sensitive to the value of $\alpha_2-\alpha_1$. Although, as already discussed, the image law is not necessarily applicable at the fields under consideration it is useful to calculate typical values of the correction based on $f(y)$ as in equation (1). These show that in general the correction is insignificant for fields below 2×10^7 v cm^{-1} but that above this value, depending on $\alpha_2-\alpha_1$, it could increase rapidly as E^{-1} decreases.

For example, the correction factor for a clean tungsten emitter where $\phi_1=4\cdot35$, $\phi_2=4\cdot65$ refer to different crystal faces increases from 0·02 at $E^{-1}=6\times10^{-8}$ to 0·3 at $E=1\cdot5\times10^{-8}$ assuming arbitrarily that $a_1=a_2$. Emitters having partial coatings such as barium on tungsten may exhibit a whole range of effective work functions and for these the effect would not be calculable. However, the important overall effect would be that the plot, figure 2, would tend to show an upcurving as the field is increased. This effect does not depend on the geometry of the emitter for its action and might quite well be observed for a planar emitter, having a constant value of E over its surface. If this effect is combined with the opposing effect due to failure of the image law then a small inflection in the observed emission curve could be expected at the highest fields in the region BC, figure 2, in agreement with experiment.

§ 4. SURFACE MICRO-GEOMETRY

Finally, consideration should be given to the other condition assumed in the derivation of equation (1), namely that the applied field E is uniform at least over distances corresponding to the barrier width at the surface, i.e. over distances of about 10 Å. The macro-geometry of the emitter–collector system produces a field enhancement which must be accurately known for interpretation of the experiments (Dyke and Trolan 1953). However, a more local field increase due to the small-scale micro-geometry, i.e. surface irregularities, may also occur. This has been discussed briefly by Benjamin and Jenkins (1940) for etched emitters. Such irregularities may be removed by flashing the emitter at a suitable temperature without a field applied (Mueller 1949). This treatment produces a definite emission pattern considered to be characteristic of a ‘smooth’ emitter. The emission is non-uniform due to the surface consisting of several crystal faces having different work functions. If these faces also have different orientations then in fact irregularities, of atomic dimensions at least, exist (Dyke and Trolan 1953, Nichols 1940, Benjamin and Jenkins 1940, Martin 1939) and it will be of some interest to estimate the effect of these on the current–field relationship.

The locally distorted field will be βE where β is an enhancement factor due to the irregularities and therefore a function of position and E is the usual value of applied field calculated from the geometry of a truly smooth system for which $\beta=1$. Equation (1) may still be used if it is considered that the resultant surface barrier has been produced by an equivalent uniform field $\bar{\beta}E$ where $\bar{\beta}$ is an average of β over the resultant barrier width. Since the barrier width decreases as E increases, in the neighbourhood of a projection $\bar{\beta}$ will also increase but for the surfaces under consideration the increase would probably be small until the barrier is confined to small distances from the emitter. Rough estimates (assuming an image law) show that for $E>4\times10^7$ v cm^{-1} the barrier width is less than 10 Å and increasing values of $\beta\,(>1)$ might be expected in this range. The modified version of equation (1) now gives

$$\ln J = \ln\left(\frac{1\cdot55\times10^{-6}E^2}{\phi}\right) + 2\ln\bar{\beta} - \frac{6\cdot85\times10^{7}\,\phi^{3/2}f(y)}{\bar{\beta}E}. \quad \ldots\ldots(3)$$

As E increases, the presence of $\bar{\beta}$ in the third term on the right-hand side of equation (3) will cause $\ln J$ to increase less rapidly than would be expected from equation (1). This effect would be partially reduced by the second term but the reduction should be small since calculations for certain ideal micro-geometries show that $\bar{\beta}$ is unlikely to exceed 2 in the present case. If it is legitimate to consider surface conditions of this sort then a further effect arises from the resulting non-uniform emission.

If r is a distance measured parallel to the macro-surface away from the axis of a given projection, i.e. into a region of smaller β, then

$$\frac{dJ}{dr} = \frac{1\cdot55\times10^{-6}}{\phi}E(2\bar{\beta}E+\alpha)\exp\left(-\frac{\alpha}{\bar{\beta}E}\right)\frac{d\bar{\beta}}{dr}. \quad \ldots\ldots(4)$$

As E increases dJ/dr will show a distinct increase since not only $\bar{\beta}$ increases but as calculations show $|d\bar{\beta}/dr|$ also increases. As a result the current will become increasingly non-uniform and estimates of current density inaccurate. If the two effects resulting from equations (3) and (4) are superimposed then the final effect may be judged from the hypothetical case shown in figure 4. Here the current densities from two adjacent regions having different values of β are shown. The resultant mean emission density, depending on relative areas as well as on β, will produce some intermediate curve as shown.

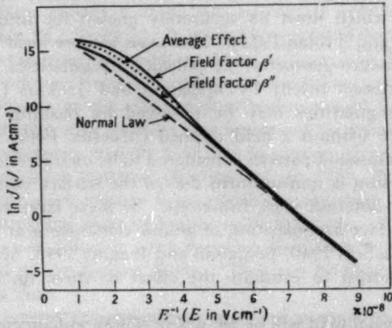


Figure 4. Emission current densities from two regions of an irregular surface having different local field factors β', β'' ($\beta' > \beta''$). The resultant average density is also shown. The deviations due to failure of the image law are not included in the figure. (Approximate scales for $\phi = 3$ ev.)

§ 5. CONCLUSIONS

It will be seen therefore that several distinct effects can be operative at fields in the range 10^7–10^8 v cm^{-1} each producing departures from the normal cold emission law. These departures fall into two categories namely those associated with the nature and uniformity of the surface barrier and those associated with distortions of the applied field near the emitter. In the first category the fact that an image-law barrier is not to be expected at small distances (~3 Å) from the surface is perhaps the most serious cause of discrepancy and would help to explain deviations from a Fowler–Nordheim law. The secondary effect due to non-uniformity of work function which is especially important for coated emitters is also not in disagreement with experiment.

In the second category, it is obvious that the emission will be sensitive to the values of the applied field in the emitter neighbourhood. The type of field distortion produced by space charges has been treated by Barbour *et al.* (1953), this being naturally related to the current rather than the field magnitude. The distortion considered in this paper is related directly to the field magnitude and provided irregularities of say atomic size exist, the resultant effects are in qualitative agreement with experiment. It is important perhaps to stress that in considering the emission patterns at fields approaching 10^8 v cm^{-1} comparatively small changes in E^{-1} require large changes in E so that, apart from any space charge effects, considerable changes should occur in those processes depending more directly on E.

ACKNOWLEDGMENTS

The author wishes to thank Professor W. J. John for facilities provided and the University of London for the award of a Turner and Newall Fellowship.

REFERENCES

BARBOUR, J. P., DOLAN, W. W., TROLAN, J. K., MARTIN, E. E., and DYKE, W. P., 1953, *Phys. Rev.*, **92,** 45.
BARDEEN, J., 1936, *Phys. Rev.*, **49,** 653.
BENJAMIN, M., and JENKINS, R. O., 1940, *Proc. Roy. Soc.* A, **176,** 262.
BURGESS, R. E., KROEMER, H., and HOUSTON, J. M., 1953, *Phys. Rev.*, **90,** 515.
DYKE, W. P., and TROLAN, J. K., 1953, *Phys. Rev.*, **89,** 799.
LANGMUIR, I., 1916, *Trans. Amer. Electrochem. Soc.*, **29,** 157.
MARTIN, S. T., 1939, *Phys. Rev.*, **56,** 947.
MUELLER, E. W., 1949, *Z. Phys.*, **126,** 642.
NICHOLS, M. H., 1940, *Phys. Rev.*, **57,** 297.
NORDHEIM, L. W., 1928, *Proc. Roy. Soc.* A, **121,** 626.
SACHS, R. G., and DEXTER, D. L., 1950, *J. Appl. Phys.*, **21,** 1304.
SLATER, J. C., 1934, *Rev. Mod. Phys.*, **6,** 209.
SOMMERFELD, A., and BETHE, H., 1933, *Handbuch der Physik*, Vol. 24, No. 2, 441 (Berlin : Springer).

Whitehead's Law of Gravitation in a Space–Time of Constant Curvature

By C. B. RAYNER
Dublin Institute for Advanced Studies [†]

Communicated by S. Raimes; MS. received 31*st January* 1955
and in amended form 10*th June* 1955

Abstract. A generalization of Whitehead's law of gravitation for a discrete set of particles having arbitrary world-lines in a flat background space–time is formulated for a space–time of constant curvature. It is then extended by a method of Synge (Rayner 1954) to include systems having finite and continuous proper-density and velocity distributions.

The new law is particularly adapted to a cosmological theory like that previously considered by the author in which every gravitating particle recedes with uniform velocity from a common spatio-temporal origin.

---

§ 1. Introduction

Whitehead's theory of gravitation requires two metrics; one is the metric of a background space–time of constant curvature, and the other defines the gravitational field. We write them

$$ds_0^2 = g_{mn}^{(0)}\, dz_m\, dz_n \qquad \ldots\ldots(1.1)$$

$$ds_g^2 = g_{mn}\, dz_m\, dz_n \qquad \ldots\ldots(1.2)$$

for any system of coordinates z_n. The signatures are $(+, -, -, -)$, Latin suffixes range (1, 2, 3, 4), and the summation convention is used.

Let $ds_{0\alpha}'$ be that (retarded) element of the world-line of the αth gravitating particle whose points[‡] lie on the past null-cones which have points of ds_0 for vertices. The basic formula in Whitehead's theory is that which expresses ds_g^2 in terms of ds_0^2 and the (discrete) distribution of matter:

$$ds_g^2 = ds_0^2 - k\sum_{\alpha} m_\alpha V_\alpha\, ds_{0\alpha}'^2, \quad (k=2G/c^2). \qquad \ldots\ldots(1.3)$$

Here the summation is over each particle P_α, of proper mass m_α, and V_α is a solution of the wave equation

$$g_{(0)}^{mn}\, V_{|mn}=0 \qquad \ldots\ldots(1.4)$$

(where the bar denotes covariant differentiation with respect to $g_{mn}^{(0)}$) subject to certain boundary conditions.

The main purpose of the present paper is to apply formula (1.3) to find the gravitational field of a discrete set of gravitating particles moving along arbitrary world-lines in a background space–time (1.1) of constant curvature. This problem was first discussed by Temple (1923), who derived by a method due to Conway (1903) an expression for the potential V, to satisfy (1.4) above. His potential function does not, however, satisfy this equation, as can be verified by direct calculation. Although we might amend his results by adding to his potential a term which he omitted, we prefer to approach the problem by a different method which appears simpler, and avoids the irrationalities in his formulae.

[†] Now at the Research Laboratories of the General Electric Company, Wembley, Middlesex.

[‡] The terms *point* and *event* both mean *point-instant* in the present paper.

§ 2. The Potential for an Isolated Event-Singularity

The line-element of a Riemannian four-space of constant, positive curvature κ, with signature appropriate to space–time, can be expressed in the Riemannian form

$$ds_0{}^2 = \frac{dx_i\,dx_i}{(1+x^2)^2} \qquad \ldots\ldots(2.1)$$

where x_i are coordinates related to real coordinates $(\bar{x}, \bar{y}, \bar{z}, \bar{t})$ by $x_1 = i\bar{x}$, $x_2 = i\bar{y}$, $x_3 = i\bar{z}$, $x_4 = c\bar{t}$, and $x^2 = \frac{1}{4}\kappa x_j x_j$.

It can be shown† that, if $R = \kappa^{-1/2}$, the invariant distance S between the arbitrary events x_i, p_j is given by

$$4R^2 \sin^2 (S/2R) = \frac{(x_i - p_i)(x_i - p_i)}{(1+x^2)(1+p^2)}. \qquad \ldots\ldots(2.2)$$

In particular, the events x_i, p_i lie on the same null geodesic if

$$\Phi \equiv (x_i - p_i)(x_i - p_i) = 0. \qquad \ldots\ldots(2.3)$$

The most general solution V_0 of the wave equation (1.4), which is a function only of the geodesic distance S_0 from the origin, is easily shown to be of the form $Af(\sigma_0) + B$, where

$$f(\sigma_0) = \frac{2\kappa\sigma_0}{1-\sigma_0{}^2} - \kappa \log \frac{1-\sigma_0}{1+\sigma_0}, \quad \sigma_0 = \cos (S_0/R). \qquad \ldots\ldots(2.4)$$

We can measure S_0 from any spatio-temporal origin p_i, and so, defining σ by

$$1 - \sigma = \frac{\frac{1}{2}\kappa(x_i - p_i)(x_i - p_i)}{(1+x^2)(1+p^2)} \qquad \ldots\ldots(2.5)$$

it follows from (2.2) that $f(\sigma)$ is a solution of (1.4).

In (2.5) let X_i be the antipode $-x_i/x^2$ of x_i; then

$$1 + \sigma = \frac{\frac{1}{2}\kappa(X_i - p_i)(X_i - p_i)}{(1+X^2)(1+p^2)} \qquad \ldots\ldots(2.6)$$

where $X^2 = \frac{1}{4}\kappa X_j X_j = 1/x^2$. From (2.4) we see that the only singularities of $f(\sigma)$ occur for $\sigma = +1$ and $\sigma = -1$. For fixed x_i (and hence, fixed X_i), the former singularities occur when p_i satisfies $\Phi = 0$, as in (2.3), and the latter when p_i satisfies

$$\Psi \equiv (X_i - p_i)(X_i - p_i) = 0. \qquad \ldots\ldots(2.7)$$

§ 3. Potential for an Arbitrary World-Line

Let $-ip_\alpha(\lambda)$ $(\alpha = 1, 2, 3)$ and $p_4(\lambda)$ be four functions (defining the world-line of a particle P) of the complex variable λ, real for real values of λ, such that the inequality

$$p_l'(\lambda) p_l'(\lambda) > 0 \qquad [p_l'(\lambda) = dp_l/d\lambda] \qquad \ldots\ldots(3.1)$$

is satisfied everywhere on the real axis; it is assumed that the four derivatives $p_l'(\lambda)$ exist on the real axis. Then from a theorem proved by Conway (1903), if the world-line $p_i(\lambda)$ does not contain the event x_i, it intersects the null-cone $\Phi(\lambda) = 0$

---

† Use coordinates $u_j = -x_j/(1-x^2)$. Then the finite translation along the u_i axis to a new origin $u_i = R \tan t_i$, $u_j = 0$ $(j \neq i, i \text{ fixed})$, is given by $\bar{u}_i = (u_i - R \tan t_i)\Delta_i$, $\bar{u}_j = (u_j \sec t_i)\Delta_i$, $\Delta_i = [1 + (u_i/R) \tan t_i]^{-1}$. Starting with $S = R \tan^{-1} [(u_r u_r)^{1/2}/R]$ for the distance S of u_r from the origin, move the origin to $u_m = p_m/(1 - \frac{1}{4}\kappa p_n p_n)$ $(m, n, r = 1, 2, 3, 4)$, and transform back to coordinates x_i. We obtain (2.2).

(equation (2.3)) in two and only two events λ_1, λ_2 and, atthese events, $\Phi(\lambda)$ has simple zeros. Moreover, one of the zeros λ_1 is such that $p_4 < x_4$, and the other λ_2 is such that $p_4 > x_4$.† Similarly, if $p_i(\lambda)$ does not contain the event X_i, it intersects the null-cone $\Psi(\lambda)=0$ (equation (2.7)) in two and only two events Λ_1, Λ_2, giving simple zeros, such that, at Λ_1, $p_4 < X_4$, and at Λ_2, $p_4 > X_4$. Accordingly $f(\sigma)$, regarded as a function of λ, has four and only four singularities, $\Lambda_1, \Lambda_2, \lambda_1, \lambda_2$ on the real axis, and at these points $d\sigma/d\lambda$ is non-zero.

Any parameter $\bar{\lambda}$, such that $d\bar{\lambda}/d\lambda$ is real and positive for all real values of λ, can be used in place of λ. Let b be an arbitrary real number; then s_0, given by

$$s_0 = s_0(\lambda) = \int_b^\lambda \frac{[p_i'(\mu)p_i'(\mu)]^{1/2}}{1+p^2(\mu)}\,d\mu, \qquad \ldots\ldots(3.2)$$

is a suitable parameter. We write $\Lambda_{01} = s_0(\Lambda_1)$, $\Lambda_{02} = s_0(\Lambda_2)$, $\lambda_{01} = s_0(\lambda_1)$, $\lambda_{02} = s_0(\lambda_2)$.

To obtain solutions of (1.4), let Γ be a contour in the complex plane of s_0 enclosing one or more of the singularities $\Lambda_{01}, \Lambda_{02}, \lambda_{01}, \lambda_{02}$. Then V, defined by

$$V = \frac{1}{2\pi i}\int_\Gamma f[\sigma\{x_1, x_2, x_3, x_4;\ p_1(s_0), p_2(s_0), p_3(s_0), p_4(s_0)\}]\,ds_0 \qquad \ldots\ldots(3.3)$$

where f is given by (2.4), and σ by (2.5), is a solution of equation (1.4).

When $\kappa=0$, Whitehead's potential function for flat space–time is obtained from (3.3) by taking for Γ a closed curve containing the one singularity λ_{01}. In this case, the logarithmic term is not present in (2.4), and the Riemann surface of $f\{\sigma(s_0)\}$ reduces to the complex plane of s_0. For κ greater than 0, however, a closed curve containing the one singularity λ_{01} is not a closed circuit on the Riemann surface of $f\{\sigma(s_0)\}$, and the resulting integral depends on the particular curve taken.

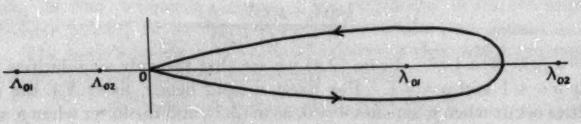


Figure 1.

We shall take for Γ the single positive loop about λ_{01}, starting from the arbitrary point $s_0=0$ (figure 1). Integrating (3.3) for this contour, we have

$$V = -\kappa\left(\frac{d\sigma}{ds_0}\right)^{-1}_{\lambda_{01}} + \kappa\lambda_{01}. \qquad \ldots\ldots(3.4)$$

Differentiate (2.5) with respect to s_0, regarding the p_j as functions of s_0 and the x_i as constants, and then put $\sigma=1$ after differentiation; then, writing $s_0'=\lambda_{01}$, it is found that

$$\left(\frac{d\sigma}{ds_0}\right)_{\lambda_{01}} = \frac{d\sigma}{ds_0'} = \frac{\kappa(x_i-p_i)(dp_i/ds_0')}{(1+x^2)\{1+p^2(s_0')\}} = \kappa W \qquad \ldots\ldots(3.5)$$

say. Hence, substituting (3.5) in (3.4), we get

$$V = -W^{-1} + \kappa s_0' \qquad \ldots\ldots(3.6)$$

---

† Since, in the coordinate system x_i, $\Phi(\lambda)$ is independent of κ, the proof is the same for flat space–time as for space–time of constant curvature.

where W is defined by (3.5), and $s_0' = s_0(\lambda_1)$, where $s_0(\lambda)$ is defined by (3.2). This is the solution of the wave equation (1.4) obtained by integrating (3.3) along the contour of figure 1; it represents the gravitational potential due to a single particle moving along an arbitrary world-line $p_i(s_0)$. When it is substituted in (1.3) the line-element of the gravitational field is obtained. Temple (1923) obtained an expression for the generalized Whitehead potential function which is equivalent to the right-hand side of (3.6), except that the term $\kappa s_0'$ is omitted. He did not take into account the contribution from the logarithmic term in (2.4) in his evaluation of the contour integral (3.3), so that his potential function is not a solution of the wave equation (1.4).

From (2.3), and the definition of ds_0' (§ 1),

$$\begin{aligned}(x_i - p_i)\,dx_i &= (x_j - p_j)(dp_j/ds_0')\,ds_0'\\ &= (1+x^2)(1+p^2)\,W\,ds_0',\end{aligned}$$

so that

$$ds_0'^2 = \frac{(x_i-p_i)(x_j-p_j)\,dx_i\,dx_j}{W^2(1+x^2)^2(1+p^2)^2} \qquad \ldots\ldots(3.7)$$

where p_j means $p_j(s_0')$.

Let the coordinates p_i in (3.6) and (3.7) refer to the αth gravitating particle. Then, substituting from (2.1), (3.6) and (3.7) in (1.3), and summing over α, we get *for the line-element of the gravitating field due to a finite number of particles of arbitrary proper mass m_α, and having arbitrary world-lines in a space–time of constant positive curvature κ,*

$$ds_g^2 = \frac{dx_i\,dx_j}{(1+x^2)^2}\left[\delta_{ij} + k\sum_\alpha \frac{m_\alpha(1-\kappa W s_0')(x_i-p_i)(x_j-p_j)}{W^3(1+p^2)^2}\right] \qquad \ldots\ldots(3.8)$$

where $x^2 = \frac{1}{4}\kappa x_j x_j$, $p^2 = \frac{1}{4}\kappa p_l p_l$. In this formula the x_j are any coordinates for which (2.1) is the metric, provided that x_1, x_2, x_3 are pure imaginary, and x_4 is real, and the p_j are the coordinates of intersection of the world-line of the αth gravitating particle with the past null-cone drawn from x_j as vertex. The quantity W is lefined by (3.5), and s_0' is the arc-length of the αth world-line measured from the rbitrary event $p_i(0)$ to the event $p_i(s_0')$. It is seen that, if κ is set equal to zero, (3.8) reduces to Whitehead's own formula in Cartesian coordinates (Rayner 954, (1.3)).†

4. THE GRAVITATIONAL FIELD DUE TO CONTINUOUS, NON-STATIC SYSTEMS IN A SPACE–TIME OF CONSTANT CURVATURE

We now extend formula (3.8) so as to include systems possessing continuous roper-density and velocity distributions. Proceeding by the method of Synge Rayner 1954, § 2), we first establish the following lemma.

Let the world-tube T intersect the null-cone

$$(x_i - p_i)(x_i - p_i) = 0 \qquad \ldots\ldots(4.1)$$

n the infinitesimal three-element $d\Sigma_0$ at p_i. Then, if $d\Sigma_n$ is the magnitude of the ction of T orthogonal to a unit covariant vector n_r, the expression

$$d\Sigma_0 = \frac{1+x^2}{1+p^2}\,\frac{d\Sigma_n}{|n_r(x_r-p_r)|} \qquad \ldots\ldots(4.2)$$

s the same value for all orientations of the unit vector n_r.

---

† Formula (3.8) is discussed in § 5.

In the world-tube T (figure 2), which has the direction of a unit covariant four-vector u_r at p_r, and normal cross section $d\Sigma_1$, the 'projection formula'

$$d\Sigma_1 = d\Sigma_n \, |g_{(0)}{}^{rs} u_r n_s| \qquad \ldots\ldots(4.3)$$

relates the magnitudes $d\Sigma_1$ and $d\Sigma_n$.

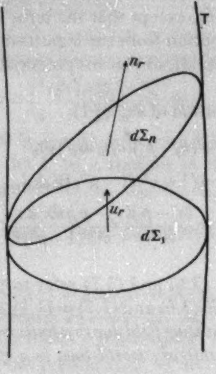


Figure 2.

Consider the pseudo-sphere of radius S and centre x_r, the surface of which consists of all points p_r such that

$$4R^2 \sin^2(S/2R) = \frac{(x_r - p_r)(x_r - p_r)}{(1+x^2)(1+p^2)}. \qquad \ldots\ldots(4.4)$$

We find, by differentiating (4.4), that the unit covariant normal V_r to the pseudo-sphere at p_r is given by

$$RV_r \sin(S/R) = -\frac{x_r - p_r + p_r(1+x^2)\sin^2(S/2R)}{(1+x^2)(1+p^2)}. \qquad \ldots\ldots(4.5)$$

Now let $d\Sigma'$ be an element of the pseudo-sphere at p_r, and let $d\Sigma_n$ be the projection of $d\Sigma'$ on an element of hypersurface drawn through p_r orthogonal to a given unit vector n_r. Then, by (4.3),

$$d\Sigma_n = d\Sigma' \, |g_{(0)}{}^{rs} n_r V_s|. \qquad \ldots\ldots(4.6)$$

Hence, defining $d\Sigma$ as the expression $d\Sigma'/S$, we have

$$d\Sigma = \frac{d\Sigma_n}{|g_{(0)}{}^{rs} n_r (SV_s)|}. \qquad \ldots\ldots(4.7)$$

Allowing S to tend to zero, and noting from (4.5) that

$$\lim_{S\to 0} (SV_s) = \frac{-(x_s - p_s)}{(1+x^2)(1+p^2)} \qquad \ldots\ldots(4.8)$$

we obtain the required result

$$d\Sigma = \frac{1+x^2}{1+p^2} \frac{d\Sigma_n}{|n_r(x_r - p_r)|}, \qquad \ldots\ldots(4.9)$$

where $d\Sigma$ is now an element of the null-cone (4.1), and $x_r - p_r$ corresponds to a nul direction at p_r. Since the expression (4.7) is independent of the orientation o

n_r, and since the limiting form of (4.7) exists, and is given by (4.9), it follows that $d\Sigma$, defined by (4.9), is independent of the orientation of n_r.

When the vector n_r is the unit covariant velocity four-vector $u_r\ (=g_{rs}^{(0)}\,dp_s/ds_0')$, $d\Sigma_n$ is the magnitude $d\Sigma_1$ of normal cross section of the world-tube T, and then

$$\begin{aligned} d\Sigma &= \frac{d\Sigma_1(1+x^2)(1+p^2)}{(x_r-p_r)(dp_r/ds_0')} \\ &= d\Sigma_1/W. \qquad \ldots\ldots(4.10) \end{aligned}$$

The number of particles contained by the three-element $d\Sigma_1$, multiplied by their proper mass, is clearly

$$\rho'\,d\Sigma_1 = \rho'\,W\,d\Sigma \qquad \ldots\ldots(4.11)$$

where ρ' is the proper density at p_r.

Hence, *the required extension of formula* (3.8) *for the line-element of the gravitational field of a distribution of matter, whose proper density* ρ' *and four-velocity* dp_r'/ds_0' *are continuous functions of position and time in a space–time of constant curvature* κ, *is*

$$ds_g^{\,2} = \frac{dx_i\,dx_j}{(1+x^2)^2}\left[\delta_{ij} + k\int\frac{(1-\kappa W s_0')(x_i-p_i)(x_j-p_j)\rho'\,d\Sigma}{W^2(1+p^2)^2}\right]. \qquad \ldots\ldots(4.12)$$

Here, $d\Sigma$ is the *absolute two-content* of the three-element at p_r on the past null-cone having x_r as vertex; it is defined by (4.9) for an arbitrary unit covariant vector n_r.

§ 5. DISCUSSION OF THE GENERALIZED LAW OF GRAVITATION

Although the results of this paper were derived under the assumption that κ is positive, they are formally the same for negative κ, since no topological properties of the background space–time in the large were assumed. If, as is usual, we make the topology definite by regarding the continuum as a four-sphere in a five-dimensional euclidean space, we find that, when κ is positive, space–time is closed in time-like and open in space-like directions, and when κ is negative, it is open in time-like and closed in space-like directions. In the former case there is an upper limit $\pi Rc^{-1}\,(R=|\kappa|^{-1/2})$ to the possible age of the universe, and in the latter case space–time has the same metrical properties as the de Sitter universe.

Perhaps the chief interest of formulae (3.8) and (4.12) lies in the presence of the factor $1-\kappa W s_0'$, which depends on the arc-length s_0' (i.e. the age) of the gravitating particle. Now, s_0', being an indefinite integral, involves an arbitrary constant, which must be defined for all particles before their gravitational field can be determined. The event $p_i(0)$ associated with this constant, which we shall call the *singular event* for the given particle, must lie in the finite history (or future, in which case s_0' will be negative) in order that s_0' may converge. It seems natural to identify the singular event with the event of creation of the particle.

Since it appears that the effect of terms in κ can only be appreciable on a cosmic scale, the results of the present paper should be considered only within the framework of a definite cosmological theory. The most appropriate application seems to be to a theory like that previously considered by the author (1954) in which every gravitating particle moves with uniform velocity away from a common spatio-temporal origin $p_i(0)$ along a geodesic in the background space–time. In such a world-model similar fundamental particles (e.g. two neutrons) would have similar gravitating properties at contemporary epochs, and the difficulty of assigning events $p_i(0)$ to individual particles would not arise. On the other hand, in a world-model in which singular events could have arbitrary temporal separation there would be a means of discriminating between otherwise identical particles, viz. by differing gravitational properties. It is hoped later to publish a short account of the properties of the world-model just indicated.

ACKNOWLEDGMENTS

This paper, written during tenure of a scholarship at the Dublin Institute for Advanced Studies, follows from part of a University of London thesis prepared at the Imperial College of Science and Technology under the supervision of Dr. G. J. Whitrow. My thanks are due to him and to Professor J. L. Synge for numerous helpful discussions during its preparation.

REFERENCES

CONWAY, A. W., 1903, *Proc. Lond. Math. Soc.*, Ser. 2, **1**, 154. (*Selected papers*, Dublin, 1953, 13–24.)

RAYNER, C. B., 1954, *Proc. Roy. Soc.* A, **222**, 509.

TEMPLE, G., 1923, *Proc. Phys. Soc.*, **36**, 176.

Thermal Vibrations of Atoms in Cubic Crystals

I. The Temperature Variation of Thermal Diffuse Scattering of X-rays by Lead Single Crystals

By L. CARTZ †

Department of Chemical Crystallography, University College, London, W.C.1

Communicated by K. Lonsdale ; *MS. received* 6*th May* 1955 *and in final form* 5*th July* 1955

Abstract. The intensity of thermal diffuse scattering of x-rays which normally increases with rise of temperature is expected to decrease with temperature increase for scattering near to high order reciprocal lattice points for crystals of low Debye characteristic temperature. Observations on lead single crystals agree well with theory. This decrease of diffuse scattering occurs when short thermal wavelengths are becoming predominant in the lattice, and so should be related to the melting point. A relationship between the Debye characteristic temperature and the melting point is obtained which has also been derived by Lindemann from other considerations.

---

§ 1. INTRODUCTION

THE intensity of thermal diffuse scattering of x-rays is known generally to increase with temperature (see table 1). However, it is expected to decrease with temperature increase for scattering near to high order reciprocal lattice points for cubic crystals of low Debye characteristic temperature Θ (Born 1942, p. 317, Laval 1943, § 9, James 1948, p. 257).

This is because the intensity of diffuse scattering D due to long thermal wavelengths modifying the cubic lattice is given by

$$D \propto T\,\mathrm{e}^{-T/\tau} = T\,\mathrm{e}^{-2M} \qquad \ldots\ldots(1)$$

for $T>\Theta$, where T is the absolute temperature, $2M$ is the Debye temperature factor, and τ is given by

$$2M = \frac{T}{\tau} = \frac{3h^2T}{mka^2\Theta^2}(h_1^2+h_2^2+h_3^2) \qquad \ldots\ldots(2)$$

where m is the mass of the atom, a the lattice parameter and $(h_1h_2h_3)$ the Miller indices of the crystal plane; τ is the 'inversion temperature' associated with the reciprocal lattice point *hkl*.

Thus D, the intensity of diffuse scattering, should increase with temperature up to $T=\tau$, and then decrease with further temperature rise.

The quantitative measurements of Laval (1939) on KCl were at temperatures below τ for the reciprocal lattice points observed and the diffuse scattering increased with temperature. Born and Lonsdale (1942) have shown that Laval's observations agree well with expression (1).

To observe the decrease of diffuse scattering with temperature rise, measurements have been carried out on lead single crystals (Lonsdale 1948 a, p. 160), lead having, as required, a low Debye characteristic temperature. Moreover, the amplitude of thermal vibrations is large, 0·28 Å (Lonsdale 1948 b),

† Now at Cavendish Laboratory, Cambridge.

so that the intensity of diffuse scattering is appreciable (Lonsdale and Smith 1942 b).

Table 1. Observations on the Effect of Temperature on Diffuse Scattering

Crystal	Temp. range of obsn (°K)	X-ray detection	Reference
Al	293–823	Photographic	Preston (1939)
NaCl	90–300	Photographic	Venkateswaran (1941)
	100–300	Ionization chamber	Baltzer 1941
	293–823	Photographic	Preston 1939
Diamond	90–290	Photographic	Lonsdale 1940 Lonsdale & Smith 1942 a
	290–600	Geiger counter	Prince 1952
	292–773	Photographic	Raman & Nilakantan 1940
	293–823	Photographic	Preston 1939
MgO	293–823	Photographic	Preston 1939
$NaNO_3$	90–290	Photographic	Lonsdale & Smith 1942 a
	90–300	Photographic	Venkateswaran 1941
	292–773	Photographic	Raman & Nilakantan 1940
α-resorcinol	90–290	Photographic	Lonsdale & Smith 1942 a
Benzil	90–290	Photographic	Lonsdale & Smith 1942 a
Sorbic acid	90–290	Photographic	Lonsdale & Smith 1942 a
KCl	90–290	Photographic	Lonsdale & Smith 1942 a
	290–665	Ionization chamber	Laval 1939 Lonsdale 1940
Lead	90–290	Photographic	Lonsdale 1942
Tin	90–290	Photographic	Bouman *et al.* 1946
p-nitroaniline	90–290	Photographic	McKeown *et al.* 1951
Quartz	290–600	Geiger counter	Prince 1952
Cu–Au alloys	290–633	Geiger counter	Cowley 1950
Pentaerithrytol	90–300	Photographic	Venkateswaran 1941
Carborundum	90–300	Photographic	Venkateswaran 1941
Fe (α)	292–585	Neutron diffraction	Lowde 1954
$CaCO_3$	93–293	Photographic	Lonsdale 1940
Pyrene	303–383	Photographic	Dutt 1953
Anthracene	90–300	Electron diffraction	Skrebowski 1948
Phloroglucine dihydrate	300–350	Photographic	Datta 1952

In all cases, an increase of diffuse scattering was observed, though only slight in the case of diamond, except that no change was observed for phloroglucine dihydrate where the temperature range studied was very small.

§ 2. THEORETICAL

A more complete expression for the intensity of diffuse scattering *D* includes a term Φ inversely proportional to the elastic constants (see Ramachandran and Wooster (1951) where the second-order diffuse scattering is also included but can be shown to be negligible for lead). Thus, in plotting *D* against *T*, allowance must be made for the temperature variation of the elastic constants and of the Debye characteristic temperature.

Siegel and Cummerow (1940) observed the Young's modulus of a single crystal of lead to decrease linearly by 11% for every 100° rise of temperature. As no other data appear to be available, all the elastic constants were assumed so to decrease (see §4), Φ thus increasing by 11% for every 100° rise in temperature.

The temperature variation of the Debye characteristic temperature given by Owen and Williams (1947), which for lead would be 2° per 100°K, was used. The actual value of Θ, however, is in some doubt (see Roberts 1951, p. 516). A value of Θ was chosen, therefore, to give the best agreement between the theoretical and the experimental temperature variation curves. This value is 78–69(±5)°K over the temperature range 150–600°K. It may be compared with the average value of Θ obtained from the temperature variation of the Bragg reflections, which is 70·3 ± 3·6°K over the temperature range 300–500°K.

§ 3. Experimental

Lead single crystals (face-centred cubic, $a = 4{\cdot}94$ Å) were grown by carefully annealing lead wires sealed into very thin soda-glass tubes; the [111] axes were found always to be parallel to the wire axes. Monochromatic Cu Kα radiation was used, the scattered x-rays being detected both photographically and by Geiger counter methods.

The diffuse scattering was observed at temperatures from 150° to 600°K. For the low temperatures, liquid air was boiled by an electric heater and the cold air stream directed on to the crystal using a Dewar tube (Reed and Lipscomb 1953). The high temperatures were obtained by surrounding the crystal on the goniometer arcs with a brass tube carrying insulation and a nichrome heater element, suitable small windows being cut. The photographic film was kept cool by air blown through jets in front of the film.

The observed x-ray intensities were corrected for Compton scattering and the mean taken of measurements from points equidistant about the reciprocal lattice points as the expression for D also contains a term depending on the distance of the reciprocal lattice point from the reciprocal lattice origin (Ramachandran and Wooster 1951, § 9).

§ 4. Results

In figures 1 and 2 are given the observations made about the $\bar{4}40$ reciprocal lattice points in the $[\bar{1}10]$ direction. At several temperatures between 161°K and

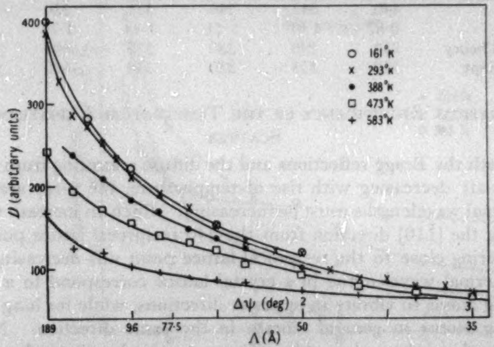


Figure 1. $\bar{4}40$ along $[\bar{1}10]$. Variation of diffuse scattering D (in abitrary units) with thermal wavelength Λ(Å) and absolute temperature. The angular settings $\Delta\psi$ corresponding to different thermal wavelengths are also given. Note the decrease of D with increase of temperature, as well as with decrease of thermal wavelength

583°K, measurements were made at different angular settings, $\Delta\psi$ away from the Bragg reflection position, where each $\Delta\psi$ corresponds to a particular thermal wavelength Λ modifying the crystal plane (Lonsdale 1942, p. 328). The calculated curve (shown dotted in figure 2) will correspond to Λ large. In figure 3 (*a*)–(*d*), observations on four other reflections are given and compared with their theoretical curves; in table 2, inversion temperatures from theory and experiment are given. All the measurements were determined using Geiger counter equipment, except where otherwise stated. The *D* scale is arbitrary in each case.

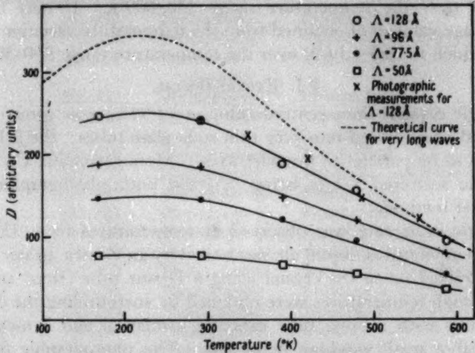


Figure 2. 440 along [$\bar{1}$10]. Variation of diffuse scattering *D* with temperature T(°K) and thermal wavelength Λ(Å). Note the maximum at about 210°K especially for long thermal waves.

Table 2

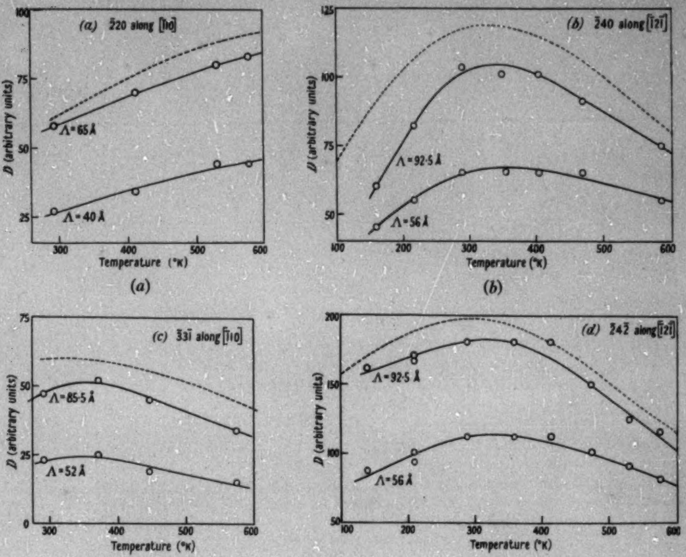
hkl		$\bar{4}40$	$\bar{2}4\bar{2}$	$\bar{2}4\bar{0}$	$\bar{3}3\bar{1}$	$\bar{2}20$	$\bar{1}3\bar{1}$
d (Å)		0·87	1·01	1·11	1·14	1·75	1·49
τ(°K)	Theory	210	290	330	330	>600	>600
	Expt	210	325	350	350	>600	>600

§ 5. THE PHYSICAL SIGNIFICANCE OF THE TEMPERATURE VARIATION OF DIFFUSE SCATTER

When both the Bragg reflections and the diffuse scattering from long thermal wavelengths are decreasing with rise of temperature, the x-ray scattering from shorter thermal wavelengths must be increasing. Such an increase was observed far out along the [$\bar{1}$10] direction from the $\bar{4}$40 reciprocal lattice point when the diffuse scattering close to the reciprocal lattice point was decreasing (figure 4).

Short thermal wavelengths in a crystal lattice correspond to a tendency of neighbouring atoms to vibrate in opposite directions, while for long wavelengths neighbouring atoms in general vibrate in the same direction. Now melting occurs when short thermal wavelengths are so predominant that any further increase of energy causes the breakdown of the lattice.

Thus we would expect the decrease of the diffuse scattering from long thermal waves to be related to the melting point.



(*a*) (*b*)

(*a*) *b*)

Figure 3. The variation of *D*, the intensity of diffuse scattering with absolute temperature for different thermal wavelengths Λ. The theoretical curves for very long waves are shown dotted. (*a*), $\bar{2}20$ along [$\bar{1}10$] ; (*b*), $\bar{2}40$ along [$\bar{1}2\bar{1}$] ; (*c*), $\bar{3}3\bar{1}$ along [$\bar{1}10$] ; (*d*), $\bar{2}4\bar{2}$ along [$\bar{1}2\bar{1}$].

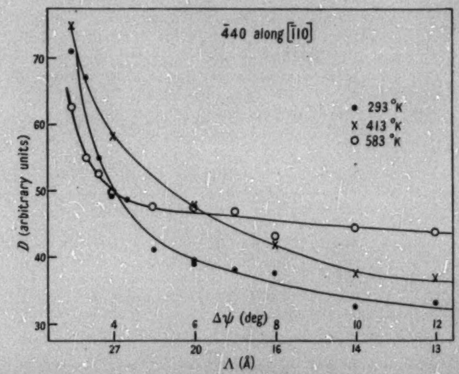


Figure 4. $\bar{4}40$ along [$\bar{1}10$]. The temperature variation of diffuse scattering *D* from short thermal wavelengths. Note that there is an increase of *D* with increase of temperature for thermal wavelengths of less than 20 Å.

Expressions (1) and (2) show that the temperature variation of D is determined by a factor $Aa^2\Theta^2[=\Omega$, say; $A=$atomic weight]. Thus, we would expect Ω to be related to the melting point T_M.

Lindemann (1910) obtained such a relationship:

$$\Omega=T_M/L=Aa^2\Theta^2 \qquad \ldots\ldots(3)$$

where L is a constant.

A discussion of this relationship and of derived properties is given in a following paper.

ACKNOWLEDGMENTS

I wish to express my thanks to Professor K. Lonsdale for considerable help whilst this work was being carried out under her careful guidance in the Department of Crystallography, University College, London.

REFERENCES

BALTZER, O. J., 1941, *Phys. Rev.*, **60**, 460.
BORN, M., 1942, *Rep. Progr. Phys.*, **9**, 294 (London : Physical Society).
BORN, M., and LONSDALE, K., 1942, *Nature, Lond.*, **150**, 490.
BOUMAN, J., ARLMAN, J. J., and VAN REIJEN, L. L., 1946, *Physica*, **12**, 353.
COWLEY, J. M., 1950, *J. Appl. Phys.*, **21**, 24.
DATTA, M. N., 1952, *Ind. J. Phys.*, **26**, 452.
DUTT, M. N., 1953, *J. Chem. Phys.*, **21**, 380.
JAMES, R. W., 1948, *The Optical Principles of the Diffraction of X-Rays* (London : Bell).
LAVAL, J., 1939, *Bull. Soc. Franç. Min.*, **62**, 137; 1943, *J. Phys. Radium*, 8, **4**, 1.
LINDEMANN, F. A., 1910, *Phys. Z.*, **11**, 609.
LONSDALE, K., 1940, *Nature, Lond.*, **146**, 806; 1942, *Proc. Phys. Soc.*, **54**, 314; 1948 a, *Crystals and X-rays* (London : Bell); 1948 b, *Acta Cryst., Camb.*, **1**, 142.
LONSDALE, K., and SMITH, H., 1942 a, *Proc. Roy. Soc.* A, **179**, 8; 1942 b, *Nature, Lond.*, **149**, 22.
LOWDE, R. D., 1954, *Proc. Roy. Soc.* A, **221**, 206.
MCKEOWN, P. J. A., UBBELOHDE, A. R., and WOODWARD, I., 1951, *Acta Cryst. Camb.*, **4**, 391.
OWEN, E. A., and WILLIAMS, R. W., 1947, *Proc. Roy. Soc.* A, **188**, 509.
PRESTON, G. D., 1939, *Proc. Roy. Soc.* A, **172**, 116.
PRINCE, E., 1952, *Thesis*, University of Cambridge.
RAMACHANDRAN, G. N., and WOOSTER, W. A., 1951, *Acta Cryst., Camb.*, **4**, 335.
RAMAN, C. V., and NILAKANTAN, P., 1940, *Proc. Indian Acad. Sci.* A, **11**, 389, 398.
REED, T. B., and LIPSCOMB, W. N., 1953, *Acta Cryst., Camb.*, **6**, 45.
ROBERTS, J. K., 1951, *Heat and Thermodynamics* (London : Blackie).
SIEGEL, S., and CUMMEROW, R., 1940, *J. Chem. Phys.*, **8**, 847.
SKREBOWSKI, J., 1948, *Proc. Phys. Soc.*, **61**, 430.
VENKATESWARAN, C. S., 1941, *Proc. Indian Acad. Sci.* A, **14**, 387.

Thermal Vibrations of Atoms in Cubic Crystals
II : The Amplitude of Atomic Vibrations

By L. CARTZ[†]
Department of Chemical Crystallography, University College, London W.C.1

Communicated by K. Lonsdale; MS. received 6th May 1955 and in final form 5th July 1955

Abstract. The Lindemann relationship between the Debye characteristic temperature and the melting point may also be obtained from a consideration of the temperature variation of thermal diffuse x-ray scattering. The amplitude of thermal vibrations of the atoms in cubic crystals can be expressed in terms of the distance between neighbouring atomic positions and the melting point. This leads to a correlation of the relative increases of disorder on melting of body-centred cubic and face-centred cubic crystals with the thermal behaviour of the elastic constants, electrical resistance, atomic diffusion, specific heat, thermal expansion and thermal conductivity. A 'law of corresponding states' may be said to exist where, at corresponding temperatures, the amplitudes of atomic vibration are the same fraction of the distance between neighbouring atomic positions in the solid. The Debye temperature factor is found to be inversely proportional to the melting point for cubic structures.

§ 1. Introduction

In a previous paper (Cartz 1955, to be referred to as I), the decrease with temperature rise, of the thermal diffuse x-ray scattering from long thermal waves modifying the crystal lattice, is related to the melting of the crystal. Thus, the temperature variation of the diffuse scattering is determined by a factor $Aa^2\Theta^2[=\Omega$, say] characteristic of the cubic crystal, where A is the atomic weight, a the lattice parameter and Θ the Debye characteristic temperature (Born 1942, Laval 1943, James 1948, p. 257). It is suggested that Ω must be dependent on T_M the melting point.

§ 2. Lindemann's Relationship

Lindemann (1910) found such a relationship. He assumed that melting occurrs when the amplitude of oscillation of the atoms becomes so great that direct collisions take place between neighbouring atoms. Then,

$$\nu^2 = \text{constant } T_M/V^{2/3}A \qquad \ldots\ldots(1)$$

where ν is the frequency of the vibrating atoms and V the atomic volume. Equating ν to the maximum frequency of Debye's theory of specific heat,

$$Aa^2\Theta^2 = T_M/L = \Omega \qquad \ldots\ldots(2)$$

where L is a constant (see also Lennard-Jones and Devonshire 1939, Blom 1913, Fürth 1944 and Born 1939).

[†] Now at Cavendish Laboratory, Cambridge.

§ 3. COMPARISON OF Ω WITH T_{M}

In tables 1 to 3, values of Ω are compared with the melting points for face-centred cubic, body-centred cubic and NaCl-type structures respectively. For the NaCl-type, the mean of the atomic weights A' is used (Lonsdale 1948). The Debye characteristic temperatures are in some doubt especially for the alkali metals. For these, two values are quoted (Blackman 1941, Mott and Jones 1936). The values of Θ for the face-centred cubic and body-centred cubic crystals are taken from James (1948) and Mott and Jones (1936), and for the NaCl-type structures from Lonsdale (1948).

Table 1. Comparison of Ω with T_{M} for Face-centred Cubic Crystals
Average $L = 14{\cdot}0\,(\pm 1{\cdot}9) \times 10^{10}$

(1)	(2)	(3)	(4)	(5)	(6) $\times 10^{10}$
A	82	39·9	5·43	80	11·1
Cu	1356	63·5	3·60	320	16·0
Ag	1233	108	4·07	210	15·7
Au	1366	197	4·06	175	13·6
Al	933	27	4·04	390	13·9
Pb	600	207	4·94	81	18·3
Ni	1725	59	3·49	400	15·0
Pd	1878	107	3·85	275	15·6
Ir	2623	193	3·82	285	11·5
Pt	2028	195	3·91	230	12·8
Rh	2218	103	3·82	370	10·8

(1) Crystal; (2) melting point T_{M} (°K); (3) atomic weight A; (4) lattice parameter a (Å); (5) Debye characteristic temperature Θ (°K); (6) $L = T_{\mathrm{M}}/Aa^2\Theta^2$.

Table 2. Comparison of Ω with T_{M} for Body-centred Cubic Crystals
Average value of L for Ta, Cr, Mo and W is $21{\cdot}6\,(\pm 2{\cdot}6) \times 10^{10}$

(1)	(2)	(3)	(4)	(5)	(6) $\times 10^{10}$
Ta	3123	181·5	3·27	245	25·9
Cr	1888	52	2·87	485	18·7
Mo	2893	96	3·14	380	21·0
W	3643	184	3·15	310	20·8
V	1983	51	3·04	300	46·7
K	335	39·1	5·20	163 100	11·9 31·7
Na	371	23	4·30	202 155	21·3 36·2
Li	459	6·94	3·50	363 400	41·0 33·8
Rb	312	85·4	5·60	85	16·2

(1) Crystal; (2) melting point T_{M} (°K); (3) atomic weight A; (4) lattice parameter a (Å); (5) Debye characteristic temperature Θ (°K); (6) $L = T_{\mathrm{M}}/Aa^2\Theta^2$.

§ 4. THE CONSTANT L

The similarity of the constant L for face-centred cubic and NaCl-type structures indicates that their dynamic properties are similar; the temperature changes in the frequency spectra of their thermal vibrations are similar when the temperature is expressed as the reduced temperature T/T_{M}.

Table 3. Comparison of Ω with T_M for NaCl-type Structures
Average value of L (neglecting LiH and PbS) is $13{\cdot}1(\pm 1{\cdot}6)\times 10^{10}$

(1)	(2)	(3)	(4)	(5)	(6) $\times 10^{10}$
NaF	1253	21·0	4·62	439	14·5
NaCl	1077	29·2	5·63	281	14·8
NaBr	1028	51·5	5·94	200	14·1
NaI	924	75·0	6·46	151	12·6
KF	1153	29·1	5·33	321	13·6
KBr	1003	59·5	6·58	177	12·4
KI	1046	83·0	7·05	132	14·4
RbCl	988	60·5	6·57	179	11·8
RbBr	955	82·7	6·87	140	12·4
RbI	915	106·2	7·33	110	13·3
CsF	956	75·9	6·01	184	10·3
LiH	953	3·98	4·10	815	21·4
PbS	1387	119·7	5·97	190	9·0

(1) Crystal; (2) melting point T_M (°K); (3) mean atomic weight A'; (4) lattice parameter a (Å); (5) Debye characteristic temperature Θ (°K); (6) $L = T_M/A'a^2\Theta^2$.

The different values of the constant L for body-centred cubic and face-centred cubic structures can be related by considering the 'open-space' of the crystal lattice. For Lindemann's relationship involves the atomic volume rather than the lattice parameter, a. By replacing a in Ω by r, half the distance between neighbouring atomic positions, which for body-centred cubic structures is $\sqrt{3}a/4$, and for face-centred cubic $a/2\sqrt{2}$, the new constant of proportionality L_r becomes the same for the two types: see table 4, where $L_r = (a/r)^2 L$.

Table 4

Crystal structure	b.c.c.	f.c.c.
$L = T_M/Aa^2\Theta^2$	$21{\cdot}6\times 10^{10}$	$14{\cdot}0\times 10^{10}$
$L_r \times T_M/Ar^2\Theta^2$	115×10^{10}	112×10^{10}

For the same structure type, different values of L will indicate different frequency spectra, i.e. differences in atomic bonding. The presence of polarization, say, in a structure may possibly be observed as a change in L. However, any variations in the value of L for the NaCl-type structures are more than accounted for by the uncertainties in the value of Θ.

The value of L for LiH is quite different, as is to be expected. Also, the alkali metals, which are relatively soft and have in general weak structures, have large values of L; that is, short thermal waves predominate even at low temperatures.

§ 5. The Amplitude of Atomic Vibrations

The mean-square amplitude of atomic vibrations $\overline{u^2}$ can be expressed in terms of the Debye characteristic temperature (James 1948, p. 220, Lonsdale 1948). For $T > \Theta$,

$$\overline{u^2} = \frac{9}{4}\,\frac{h^2 T}{\pi^2 m k \Theta^2} = 4{\cdot}364\times 10^{-14}\,\frac{T}{A\Theta^2} \qquad \ldots\ldots(3)$$

where m is the actual mass of the atom.

Using the Lindemann relation, we have for $T > \Theta$

$$\frac{(\overline{u^2})^{1/2}}{a} = \left(4{\cdot}364 \times 10^{-14}\,\frac{LT}{T_{\mathrm{M}}}\right)^{1/2}. \qquad \ldots\ldots(4)$$

Thus, the root mean square amplitude of atomic vibrations depends on the lattice parameter, open-space and melting point of the lattice.

At the melting point, $(\overline{u^2})_{\mathrm{M}}{}^{1/2}$ is a certain fraction of the lattice parameter: 9·6% for body-centred cubic, and 7·8% for face-centred cubic and NaCl-type structures. However, these values are only approximate as (4) may no longer be valid near the melting point, see § 8.

Expression (4) can be rewritten, for face-centred cubic and body-centred cubic structures, for which $L = (r/a)^2 L_r$, where $L_r = 114 \times 10^{10}$,

as
$$(\overline{u^2})^{1/2} = 0{\cdot}22\, r \left(\frac{T}{T_{\mathrm{M}}}\right)^{1/2} \qquad \ldots\ldots(5)$$
or
$$(\overline{u^2})_{\mathrm{M}}{}^{1/2} = 0{\cdot}22\, r.$$

This corresponds to the assumption made by Lindemann (1910) that ρ is a constant for all substances, where $\sigma = r(1-\rho)$, σ being the 'sphere of action' of the atoms, so that, as neighbouring atoms are assumed to collide at the melting point,

$$r\rho = (\overline{u^2})_{\mathrm{M}}{}^{1/2}. \qquad \ldots\ldots(5a)$$

Thus the atomic radii calculated by the method of Landé need a correction to allow for the thermal vibrations of the atoms.

Grüneisen (1926) has shown that

$$\frac{(\overline{u^2})^{1/2}}{2r} = \left(\frac{V - V_0}{8\pi^2 \gamma z_0{}^2 V_0/3}\right)^{1/2} \qquad \ldots\ldots(6)$$

where V, V_0 are the atomic volumes at temperatures of T and 0°K respectively, γ is Grüneisen's constant, z_0 is a function, given by Grüneisen, of n and m, where the potential energy of the atoms in the solid is given by $-a/r^m + b/r^n$, a and b being constants.

Grüneisen showed that $(V - V_0)/V_0 \simeq 0{\cdot}08$ at the melting point, and $8\pi^2\gamma z_0{}^2/3 \simeq 8$ for monatomic metals, so that $(\overline{u^2})_{\mathrm{M}}{}^{1/2}$ is 20% of r at the melting point. This may be compared with $(\overline{u^2})_{\mathrm{M}}{}^{1/2} = 0{\cdot}22r$ from (5). Again $(\overline{u^2})_{\mathrm{M}}{}^{1/2}$ can be estimated from (5a). The value of ρ is given by Roberts (1951, p. 552)

$$\sqrt{2}\boldsymbol{R}^{1/2}\boldsymbol{N}^{1/2}/\pi\rho = C = 2{\cdot}06 \times 10^{12}$$

where $\boldsymbol{R}$ is the gas constant, $\boldsymbol{N}$ Avogadro's number and C a constant determined by comparing the observed Einstein atomic frequencies with the Lindemann relationship. Then $\rho = (\overline{u^2})_{\mathrm{M}}{}^{1/2}/r = 0{\cdot}17$. Also, from § 7·3, using $b = 1{\cdot}86$, $(\overline{u^2})_{\mathrm{M}}{}^{1/2}/r = 0{\cdot}18$.

In table 5, values of $(\overline{u^2})^{1/2}$ at room temperatures calculated from (4) or (5) are compared with previous values based on Θ (see also Lonsdale 1948).

§ 6. COMPARISON OF VIBRATIONS IN FACE-CENTRED CUBIC AND BODY-CENTRED CUBIC STRUCTURES

Consider a cube of side a into which several atoms can be placed in a face-centred cubic arrangement, or fewer but larger atoms can be placed in a body-centred cubic arrangement. The larger value of L for the body-centred cubic arrangement means that short thermal wavelengths predominate at a lower temperature relative to the melting point. Again, since the amplitude of vibration is proportional to r, it will be greater in the body-centred cubic than in the face-centred cubic case. At the same temperature relative to the melting point, the vibrations of the body-centred cubic structure are of greater amplitude and have a greater predominance of short thermal waves.

Table 5. Root mean square amplitudes of atomic vibrations at room temperatures: comparison of values derived from expressions (4) or (5) and from Θ. For the NaCl-type structures, the previous values of $(\overline{u^2})^{1/2}$ are taken from Lonsdale (1948).

f.c.c.	$(\overline{u^2})^{1/2}$ (Å)		b.c.c.	$(\overline{u^2})^{1/2}$ (Å)		NaCl-type	$(\overline{u^2})^{1/2}$ (Å)	
	From (5)	Previous values		From (5)	Previous values		From (5)	Previous values
Pb	0·27	0·32	W	0·085	0·085	NaF	0·17	0·19
Al	0·17	0·18	Ta	0·095	0·11	KF	0·21	0·21
Ag	0·15	0·16	Mo	0·10	0·10	NaCl	0·22	0·24
Cu	0·13	0·15	V	0·11	0·16	KI	0·29	0·31
Au	0·15	0·15	Cr	0·11	0·11	NaBr	0·25	0·25
Pd	0·12	0·13	Li	0·27	0·37	KBr	0·28	0·27
Pt	0·11	0·12	Na	0·37	0·50	RbCl	0·28	0·26
Ir	0·10	0·09	K	0·47	0·57	CsF	0·26	0·23
Ni	0·11	0·12	Rb	0·52	0·45	RbBr	0·29	0·29
Rh	0·11	0·10				NaI	0·28	0·28
						RbI	0·32	0·32
						LiH	0·18	0·24
						PbS	0·21	0·18

Thus, on melting, the magnitude of the increase of disorder should be smaller for body-centred cubic structures. In table 6, the latent heat of fusion, the change of entropy, electrical resistance and of volume on melting are recorded. The smaller increase is for the body-centred cubic structures.

Table 6. Comparison of the Increase of Disorder on Melting of Body-centred Cubic and Face-centred Cubic Crystals†

(1)	(2)	(3)	(4)	(5)
Body-centred cubic				
Li	3440	1·85	1·68	1·5
Na	2500	1·7	1·45	2·5
K	2300	1·7	1·55	2·5
Rb	2180	1·7	1·61	2·5
Cs	2025	1·7	1·66	2·5
Face-centred cubic				
Cu	11300	2·0	2·1	4·1
Ag	11300	2·1	1·9	4·5
Au	12900	2·4	2·3	5·2
Al	10500	2·56	1·64	6·7
Pb	5460	2·0	2·1	3·4

(1) Crystal type; (2) latent heat of fusion (joule/g mol); (3) change of entropy on melting; (4) ratio of electrical resistance in liquid to that in solid; (5) volume change on fusion ($100\Delta V/V_s$).

† Roberts, 1951, p. 254, Mott and Jones, 1936, p. 278, *International Critical Tables*, 1926, **2**, 474.

§ 7. The Temperature Variation of the Amplitude of Thermal Vibrations

Expression (4) shows that, at temperatures above Θ, the variation with temperature of $(\overline{u^2})^{1/2}$, relative to the lattice parameter, is the same for the cubic substances provided that the temperature is expressed as the reduced temperature T/T_M. That is, that

$$\overline{u^2}/(\overline{u^2})_M = T/T_M \qquad \ldots\ldots(7)$$

where $(\overline{u^2})_M$ is the mean square amplitude of vibration at the melting point.

Thus, properties which depend on the vibrations of the atoms, such as the elastic constants, electrical resistance, atomic diffusion, specific heat, thermal expansion and thermal conductivity, should also vary similarly with reduced temperature.

In the following sections, the temperature variations of the thermal properties ξ are expressed as

$$\frac{\xi}{(\xi)_{M/2}} = f\left(\frac{T}{T_M}\right). \qquad \ldots\ldots(8)$$

A linear relationship is obtained in all cases over a temperature range approximately given by $\sim 0{\cdot}2 < T/T_M < \sim 0{\cdot}7$; in general, the condition that $T > \Theta$ is satisfied (see § 8). The value of $(\xi)_{M/2}$, i.e. at $T = \frac{1}{2}T_M = T_{M/2}$ has been used, as the value at the melting point may be influenced by other factors (see § 8).

7.1. *The Elastic Constants*

Sutherland (1891) obtained such a relationship for the temperature variation of Young's modulus;

$$E/E_M = 1 - \eta\,(T/T_M) \qquad \ldots\ldots(9)$$

where E, E_M are the Young's moduli at T and T_M °K respectively, and η is a constant for all metals. Siegel and Cummerow (1940) have noted that the Young's modulus of Pb, the rigidity modulus of Al, and c_{44} for Na all vary similarly with the reduced temperature.

7.2. *The Electrical Resistance*

The electrical resistance R is proportional to $\overline{u^2}$ (Mott and Jones 1936, p. 243). On plotting the relative change in R against T/T_M for face-centred cubic and body-centred cubic metals one straight line is obtained (see figure 1) and is given by

$$\frac{R}{R_{M/2}} = 2{\cdot}09\left(\frac{T}{T_M}\right) - 0{\cdot}045 \qquad \ldots\ldots(10)$$

for the range $\sim 0{\cdot}1 < T/T_M < \sim 0{\cdot}6$. The data are taken from the *International Critical Tables*, 1926, **6**, 136, where they are presented in the form

$$r_t = r_{t'}[1 + \alpha(t - t') + \beta(t - t')^2] \qquad \ldots\ldots(11)$$

r_t being the specific resistance at t°C and α and β the first and second order temperature coefficients of resistance: β has been neglected in calculating figure 1, which is justifiable up to a temperature $T/T_M \simeq 0{\cdot}6$.

Comparing (10) and (11), neglecting β and rearranging, we obtain

$$\alpha = \frac{2{\cdot}09}{T_M - 2{\cdot}09(T_{M/2} - T')} \qquad \ldots\ldots(12)$$

where $T' = t' + 273$.

The alkali metals have been excluded from figure 1 as data are only available for them near to their melting points where β can no longer be neglected.

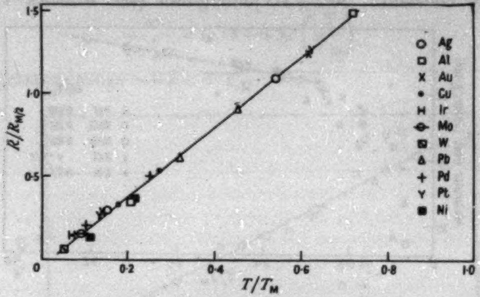


Figure 1. The variation of electrical resistance with reduced temperature.

7.3. *Atomic Diffusion*

Atomic diffusion in solids is observed to vary with temperature as given by

$$D = B \exp(-E/\boldsymbol{R}T) \qquad \ldots\ldots(13)$$

where D is the diffusion coefficient and B and E are constants for each diffusion system (Mehl 1937). On plotting $\log D$ against $1/T$ straight lines are obtained of different slopes E for the different diffusion systems. Assuming that diffusion depends on the vibrations of the atoms, then if D is plotted against T_M/T, the inverse of the reduced temperature, the straight lines for the different diffusion systems should all have the same slope of value E/T_M. This has been observed; table 7 taken from Barrer (1940) shows that E/T_M is a constant. Mehl (1937) noted that at a given temperature, the rate of self-diffusion is inversely related to the melting point.

Table 7

System	Cu in Au	Cu in Ni	Mo in W
T_M for solvent metal (°K)	1356	1728	3473
E/T_M	20·2	20·6	21·3

From Barrer (1940).

Van Liempt (1932) incorporated Lindemann's theory of melting with the temperature dependence of diffusion and so obtained $E/T_M = 3\boldsymbol{R}b^2$ where $b = 2r\rho/(\overline{u^2})_M{}^{1/2}$ and is about 2. Using the average value of $E/T_M = 20{\cdot}7$ from table 7, $b = 1{\cdot}86$. This gives another estimate of $(\overline{u^2})_M{}^{1/2}/r$, see § 5.

7.4. *The Atomic Heat*

In figures 2 and 3, the atomic (or molecular) heat at constant pressure C_p is plotted against the reduced temperature for NaCl-type structures and for cubic metals.

For the NaCl-type structures, over the temperature range

$$\sim 0{\cdot}35 < T/T_M < \sim 0{\cdot}8,$$

a straight-line relationship is obtained, given by

$$C_p = 48{\cdot}6 + 11{\cdot}8(T/T_M) \text{ joule/g mol.} \qquad \ldots\ldots(14)$$

Thus, in table 8, the molecular heats at $T_{M/2}$ for the alkali halides are seen to be constant at 54·5 joule/g mol; however, for the cubic metals, $(C_p)_{M/2}$ is only approximately constant: $30{\cdot}2 \pm 3{\cdot}3$ joule/g mol (see Roberts 1951, p. 288).

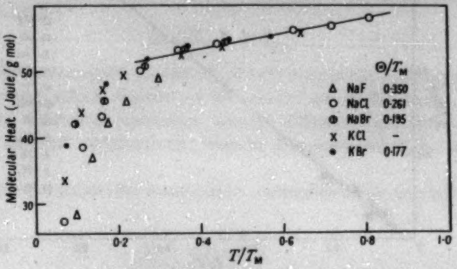


Figure 2. The variation of molecular heat of NaCl-type structures with reduced temperature.

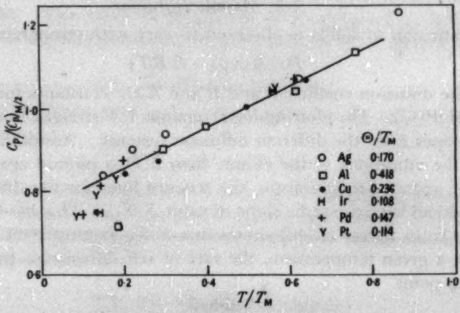


Figure 3. The variation of atomic heat of face-centred cubic metals with reduced temperature.

In figure 3 a straight-line relationship for the metals is found, over the temperature range $\sim 0{\cdot}25 < T/T_M < \sim 0{\cdot}7$, given by

$$\frac{C_p}{(C_p)_{M/2}} = 0{\cdot}76 + 0{\cdot}48\left(\frac{T}{T_M}\right). \qquad \ldots\ldots(15)$$

The data for the metals are taken from Grüneisen (1908), and for the NaCl-type structures from the *International Critical Tables*, 1926, **5**, 95.

Table 8

(1)	(2)	(3)	(4)
NaCl	54·5	Ag	28·0
NaBr	55·0	Al	24·0
KCl	54·0	Cu	28·6
KBr	54·8	Ir	35·4
		Pd	31·8
		Pt	32·6
Mean value	$54{\cdot}5 \pm 0{\cdot}4$	Mean value	$30{\cdot}2 \pm 3{\cdot}3$

(1) NaCl-type structure; (2) molecular heat at $T_{M/2}$ (joule/g mol); (3) cubic metal; (4) atomic heat at $T_{M/2}$ (joule/g atom).

7.5. *The Thermal Expansion*

The coefficient of thermal expansion α, at various temperatures, for several cubic metals is shown in figure 4 to be given by

$$\frac{\alpha}{(\alpha)_{\mathrm{M}/2}}=0\cdot74+0\cdot52\left(\frac{T}{T_{\mathrm{M}}}\right) \qquad \ldots\ldots(16)$$

over the temperature range $\sim0\cdot2<T/T_{\mathrm{M}}<\sim0\cdot7$.

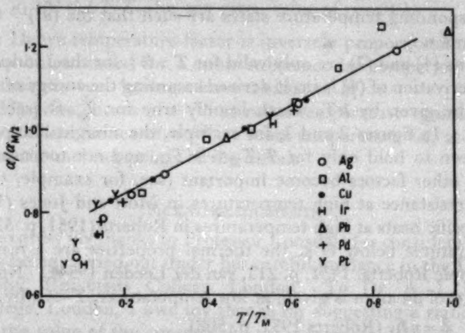


Figure 4. The variation of the coefficient of thermal expansion with reduced temperature.

The data for Cu, Pd, Ag, Ir and Pt are taken from Roberts (1951, p. 260), for Pb from Stokes and Wilson (1941) and for Al from Wilson (1941).

From (16) and (5) it is seen that

$$\frac{\alpha}{(\alpha)_{\mathrm{M}/2}}=0\cdot74+10\cdot7\left(\frac{\overline{u^2}}{r^2}\right). \qquad \ldots\ldots(17)$$

Grüneisen (1912) has shown that the volume coefficient of expansion is given by by $(\overline{u^2}/r^2)$; see (6). Grüneisen (1910) has compared the reduced temperature with the expansion δl (where if l is the length at 0°K, $l+\delta l$ is the length at T°K), for Ni, Pd, Ir, Al, Pb and Bi, from $T/T_{\mathrm{M}}=0$ to 1·0. The slopes of the curves, which equal $T_{\mathrm{M}}dl/dt$, i.e. the melting point times the coefficient of thermal expansion, are in general not the same for the different metals, though they are the same for Pd, Ir and Al from $T/T_{\mathrm{M}}=0\cdot3$ to 0·7. Grüneisen has concluded, therefore, that at any given value of the reduced temperature, the relationship $T_{\mathrm{M}}dl/dt=$ constant is not generally true (see also Blom 1913, Zwikker 1954).

7.6. *Thermal Conductivities*

The thermal conductivities k, at various temperatures, of Cu, Ag and Pb are observed to lie on the same straight line (figure 5) given by

$$\frac{k}{k_{\mathrm{M}/2}}=1\cdot116-0\cdot232\left(\frac{T}{T_{\mathrm{M}}}\right) \qquad \ldots\ldots(18)$$

though the values for Al diverge considerably from this line.

The data for Ag, Pb and Al are taken from the *International Critical Tables*, 1926, **5**, 220, and for Cu from Hume-Rothery (1931).

§ 8. Law of Corresponding States

Fürth (1944) has shown that

$$T_M = \text{constant} \times T_c \qquad \ldots\ldots(19)$$

where T_c is the critical temperature of the solid. It follows from the data presented in § 7 that a 'law of corresponding states' holds for these cubic solids, where (7) is the 'reduced equation of state'. The comparison of thermal properties on a temperature scale T/T_M is a comparison on a basis of $\overline{u^2}/r^2$ (see equation (5)); that is, corresponding temperature states are such that the $(\overline{u^2})^{1/2}$ values are the same fraction of r.

Expressions (4) and (7) are only valid for $T > \Theta$; for the Lindemann relation used in the derivation of (4) is itself derived assuming the energy of a single linear oscillator to be given by $\boldsymbol{k}T$, which is only true for $T > \Theta$ (see also paper I, equation (1)). In figures 2 and 3, for example, the straight line portions of the graphs are seen to hold only for $T/T_M > \Theta/T_M$, and not too near the melting point, where other factors become important (see, for example, the discussion on electrical resistance at high temperatures in Mott and Jones (1936, p. 269), or that on specific heats at high temperatures in Roberts (1951, p. 522)).

At temperatures below Θ°K, the thermal properties are a function of T/Θ (see, for example Roberts, 1951, p. 213, van der Leeden 1954). Now, the energy of oscillation ψ of an atom is given, at low temperatures, $T < \Theta$, by

$$\psi = n\boldsymbol{h}\nu \text{ (Roberts 1951, p. 508)}$$
$$= \text{constant}.\ n\Theta \text{ where } n \text{ depends on the temperature}$$

and at higher temperatures, $T > \Theta$, by

$$\psi = 2\pi^2\nu^2 m\overline{u^2}$$
$$= \text{constant}.\ m\Theta^2\overline{u^2}$$
$$= \text{constant}.\ n'm\Theta^2 r^2$$

since $\overline{u^2} = n'r^2$, where n' depends on the temperature. So that, using (2)

$$\psi = \text{constant}.\ n'T_M. \qquad \ldots\ldots(20)$$

That is, over the whole temperature range, except near the melting point, the thermal properties are a function of T/ψ.

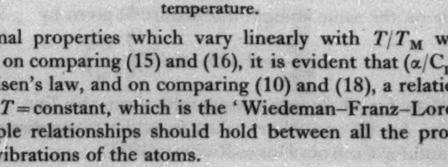


Figure 5. The variation of thermal conductivity of face-centred cubic metals with reduced temperature.

Those thermal properties which vary linearly with T/T_M will be simply related. Thus, on comparing (15) and (16), it is evident that $(\alpha/C_p)_T \simeq$ constant, which is Grüneisen's law, and on comparing (10) and (18), a relation is obtained of the form $kR/T =$ constant, which is the 'Wiedeman–Franz–Lorenz law'.

Similar simple relationships should hold between all the properties which depend on the vibrations of the atoms.

§ 9. THE DEBYE TEMPERATURE FACTOR

The Debye temperature factor M is given for cubic substances by

$$2M = \frac{3\boldsymbol{h}^2T}{mka^2\Theta^2}(h_1^2 + h_2^2 + h_3^2) \qquad \ldots\ldots(21)$$

where $(h_1h_2h_3)$ are the Miller indices of the reflection.

Using (2), we obtain

$$M \propto LT/T_M \text{ for any given value of } (\sin\theta)/\lambda. \qquad \ldots\ldots(22)$$

Thus, the Debye temperature factor is inversely proportional to the melting point and may be estimated for a cubic substance by comparing with another of the same structure-type having a known value of M.

Blake (1933) gives curves of e^{-2M} at room temperature against $(\sin\theta)/\lambda$ for many metals. These values of e^{-2M} at room temperature lie, at any $(\sin\theta)/\lambda$, in the order of the melting points of the metals.

ACKNOWLEDGMENTS

I wish to express my thanks to Professor Lonsdale for considerable help while this work was being carried out under her careful guidance in the Department of Crystallography, University College, London. To Dr. A. J. E. Welch, of Imperial College, London, I owe my thanks for suggesting a significance in the variations in the value of the constant L.

REFERENCES

BARRER, R. M., 1940, *Proc. Phys. Soc.*, **52,** 58.
BLACKMAN, M., 1941, *Rep. Progr. Phys.*, **8,** 11 (London : Physical Society).
BLAKE, F. C., 1933, *Rev. Mod. Phys.*, **5,** 169.
BLOM, C. E., 1913, *Ann. Phys., Lpz.*, **42,** 1397.
BORN, M., 1939, *J. Chem. Phys.*, **7,** 591; 1942, *Rep. Progr. Phys.*, **9,** 294 (London: Physical Society).
CARTZ, L., 1955, *Proc. Phys. Soc.* B, **68,** 951.
FÜRTH, R., 1944, *Proc. Roy. Soc.* A, **183,** 87.
GRÜNEISEN, E., 1908, *Ann. Phys., Lpz.*, **26,** 211; 1910, *Ibid.*, **33,** 33; 1926, *Handb. Phys.*, **10,** 50.
HUME-ROTHERY, W., 1931, *The Metallic State* (Oxford : University Press).
International Critical Tables, 1926 (New York : McGraw-Hill).
JAMES, R. W., 1948, *The Optical Principles of the Diffraction of X-rays* (London : Bell).
LAVAL, J., 1943, *J. Phys. Radium*, (8), **4,** 1,
VAN DER LEEDEN, 1954, *Proc. Kon. Ned. Ak. v. Wet.*, Ser. B, **57,** No. 3, 317.
LENNARD-JONES, J. E., and DEVONSHIRE, A. F., 1939, *Proc. Roy. Soc.* A, **170,** 464.
VAN LIEMPT, J. A. M., 1932, *Rec. Trav. Chim. Pays-Bas*, **51,** 114.
LINDEMANN, F. A., 1910, *Phys. Z.*, **11,** 609.
LONSDALE, K., 1948, *Acta Cryst., Camb.*, **1,** 142.
MEHL, R. F., 1937, *J. Appl. Phys.*, **8,** 174.
MOTT, N. F., and JONES, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford : University Press).
ROBERTS, J. K., 1951, *Heat and Thermodynamics* (London : Blackie).
SIEGEL, S., and CUMMEROW, R., 1940, *J. Chem. Phys.*, **8,** 847.
STOKES, A. R., and WILSON, A. J. C., 1941, *Proc. Phys. Soc.*, **53,** 658.
SUTHERLAND, W., 1891, *Phil. Mag.*, **32,** 31, 215, 524.
WILSON, A. J. C., 1941, *Proc. Phys. Soc.*, **53,** 235.
ZWIKKER, C., 1954, *Physical Properties of Solid Materials* (London : Pergamon Press), p. 158.

RESEARCH NOTES

The Effect of Illumination on the Excess Noise of Silicon Filaments

By C. A. HOGARTH and W. BARDSLEY
Radar Research Establishment, St. Andrew's Road, Great Malvern, Worcs.

MS. received 18*th April* 1955

In a recent publication by Lawrance, Gibson and Granville (1954) the properties of germanium filamentary transistors were described. At low temperatures experimental results indicated that trapping effects were giving rise to high values of current gain α at low values of emitter current. When operated in this temperature range certain analogies between the current noise in semiconductor filaments and the α values of filamentary transistors were drawn. From these the suggestion was made that since traps in a filamentary specimen give rise to increased α values, they should also cause the assembly to act as a 'built-in amplifier' for excess noise particularly if, as Montgomery (1952) suggests, this noise is associated with minority carrier generation. Thus the reduction of trapping, for example by illumination, should reduce the excess noise power. It is known (Haynes and Hornbeck 1953) that there is considerable trapping in p-type silicon crystals at room temperature and rather less in n-type samples. Experiments have therefore been made to test the prediction that the excess noise in silicon filaments is in fact reduced by irradiation with white light.

We have grown high-purity silicon crystals for experiments on transistor assemblies and have measured the excess noise power for several filamentary samples using a simple measuring circuit and an amplifier tuned to 5 c/s with a bandwidth of 0·45 c/s. The filaments could be illuminated by means of a 240 w d.c. tungsten lamp, using a 1-inch slab of Perspex as an absorber for the heat radiation. The current I_b through the specimens was taken as a measuring parameter and it was verified that I_b was proportional to the bias voltage. An estimate of the relative light intensity I was obtained by taking I proportional to $1/x^2$ where x was the distance of the lamp from the specimen.

Figure 1 shows the effect of illumination on two representative samples of 30 ohm cm silicon (*a*) p-type and (*b*) n-type. In each case there is a reduction in noise on illumination. This reduction is greater for smaller sweeping currents but is rather less in the case of n-type than in p-type material of comparable purity and for specimens which had been subjected to similar treatments such as etching and washing.

Figure 2 shows how the decrease in noise power for the p-type specimen varies with light intensity. The decrease in Johnson noise due to the increased conductance of the specimen under illumination was calculated for each level of illumination. In every case the decrease of noise power, even for the large sweeping currents, is greater than the appropriate decrease in Johnson noise.

The interpretation of the above results is that the free minority carriers produced by the absorbed light keep the traps occupied and thus reduce their

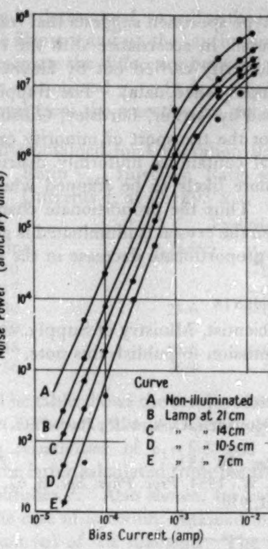


Figure 1 (*a*). Noise power under varying conditions of illumination as a function of current for p-type silicon (30 ohm cm).

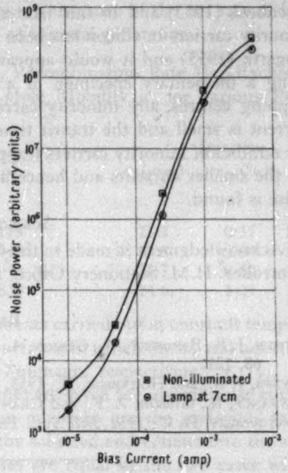


Figure 1 (*b*). Noise power as a function of current for a filament of 30 ohm cm n-type silicon with and without background illumination.

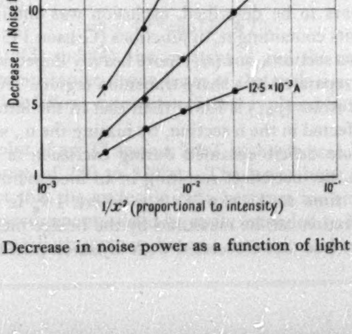


Figure 2. Decrease in noise power as a function of light intensity.

effectiveness. The greater effect for the p-type specimen suggests that trapping is initially more pronounced in p-type material, in accordance with the results of experiments on the decay of photoconductivity carried out by Haynes and Hornbeck (1953) and in this laboratory (unpublished data). The trapping of minority carriers in silicon has been discussed by Arthur, Bardsley, Gibson and Hogarth (1955) and it would appear that for the transport of minority carriers along a filamentary specimen of a material containing uniformly distributed trapping centres, any minority carrier is more likely to be trapped when the current is small and the transit time long. Thus the proportionate change in the number of minority carriers trapped when the crystal is illuminated is greater for the smaller currents and hence a larger proportionate decrease in the excess noise is found.

ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this note.

REFERENCES

ARTHUR, J. B., BARDSLEY, W., GIBSON, A. F., and HOGARTH, C. A., 1955, *Proc. Phys. Soc.* B, **68**, 121.

HAYNES, J. R., and HORNBECK, H., 1953, *Phys. Rev.*, **90**, 152.

LAWRANCE, R., GIBSON, A. F., and GRANVILLE, J. W., 1954, *Proc. Phys. Soc.* B, **67**, 625.

MONTGOMERY, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.

Some Measurements connected with Carrier Deficit Phenomena in Germanium

BY A. MANY, D. GERLICH AND E. HARNIK
Department of Physics, The Hebrew University, Jerusalem, Israel

MS. received 25*th April* 1955 *and in amended form* 14*th June* 1955

CARRIER deficit phenomena in transistor materials have been reported by several authors (Banbury 1953, Many 1954, Angello and Ebert 1954, Gibson 1954, Bray, private communication). These can be classified, following Low (1955), into two categories, namely extraction and exclusion. The present communication is concerned with the latter type of carrier concentration disturbance, where the composition ratio ($I_{\text{minority}}/I_{\text{total}}$) of the current entering the filament is lower than the normal equilibrium ratio in the bulk. In the measurements to be described, exclusion was obtained by using n-type germanium filaments containing n_+/n junctions (Gibson 1954). These filaments are composed of two sections, one (n_+) more heavily doped with donor impurities than the other (n), separated by a sharp transition region. As the current composition ratio in the former (γ_{n+}) is lower than that in the latter (γ_n), carrier (hole) exclusion can be effected in the n section, by biasing the n_+ section in the forward direction. The hole deficit obtained during exclusion is partly recovered by thermal generation, the net effect resulting in an increase of filament resistance, which varies with time as $1-\exp(-t/\tau_g)$, where $1/\tau_g$ is the generation rate. This resistivity variation can be measured by the bridge method (Many 1954) in an analogous manner to that associated with injection phenomena. In this way

τ_g can be determined, as well as the transit time of the deficit through the n section of the filament and γ_{ex} the 'exclusion ratio'. The latter expression is introduced to designate the difference between the composition ratio of the current entering the n section and γ_n. In fact this quantity is a measure of the net rate of hole removal from the n section, in the absence of thermal generation.

Results obtained for Generation Rate $1/\tau_g$, Recombination Rate $1/\tau_r$, Exclusion Ratio γ_{ex}, Injection Ratio γ_{in} and Hole Mobility μ_p for various Pulse Currents I, at $T = 293°$K

	Exclusion			Injection		
I (mA)	τ_g (μsec)	γ_{ex} (%)	μ_p	τ_r (μsec)	γ_{in} (%)	μ_p
10		−0·81				
17		−0·80				
30	25·6	−0·88	2115	23·5	0·37	1900
40	25·6	−0·84	2040	23·5	0·64	1880
48	25·6	−0·82	2040	23·7	0·80	1870
56	25·4	−0·85	1830	24·6	1·31	1800

The table shows the results of measurements carried out at constant temperature (293°K) on a filament with roughened surfaces, in which the n_+ and n sections have resistivities of $\rho_1 = 5{\cdot}3$ and $\rho_2 = 13{\cdot}7$ ohm cm, respectively. τ_g, γ_{ex} and μ_p (the latter calculated from the transit time) are given for several pulse current amplitudes I. Also shown, for comparison purposes, are the analogous values in the case of injection, obtained through the soldered end-contact into the same section (n) of the filament. The mobilities are equal in the two cases within the accuracy of transit time measurements. Although the recombination rate $1/\tau_r$ is consistently greater than the corresponding generation rate, the difference might well be within the limit of error. In contrast with injection ratio γ_{in}, which increases linearly with current, the exclusion ratio is constant, independent of the applied voltage. Moreover, this constant is found to be equal, within the experimental error, to the calculated value of $\gamma_{n+} - \gamma_n$, at the same temperature, as is to be expected on theoretical grounds. The relation

$$\gamma_{ex} = \gamma_{n+} - \gamma_n \qquad \ldots\ldots(1)$$

was further confirmed, in a more general manner, as follows.

$$\gamma_{n+} - \gamma_n = p_1/(p_1 + bn_1) - p_2/(p_2 + bn_2) = e\mu_p(\rho_1 p_1 - \rho_2 p_2) \qquad \ldots\ldots(2)$$

where the suffixes 1 and 2 refer to the n_+ and n sections, respectively. Assuming, for simplicity, an extrinsic range of temperatures ($p \ll n$), equation (2) leads to

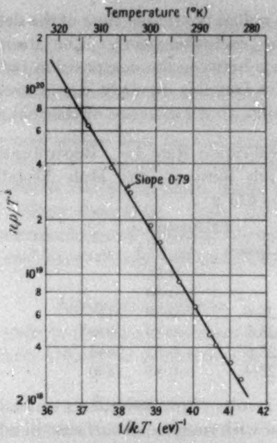
$$\gamma_{n+} - \gamma_n = e^2\mu_p\mu_n(\rho_1^2 - \rho_2^2)np \qquad \ldots\ldots(3)$$

where (Shockley 1950)

$$np = AT^3 \exp(E_g/\boldsymbol{k}T). \qquad \ldots\ldots(4)$$

Using equations (1), (3) and (4), the constant A and the energy gap E_g can be deduced from the temperature dependence of the exclusion ratio. Actually, in the measurements that follow, expression (3) was replaced by the general formula, not assuming $p \ll n$.

In the figure a plot of np/T^3 against $1/\boldsymbol{k}T$ is shown on a logarithmic scale, in the temperature range of 280 to 320°K. The lower limit of temperature was imposed by the fact that the exclusion ratio becomes too small to be accurately measured. The measurements could be easily extended to higher temperatures, up to the range where both sections become intrinsic and exclusion disappears.



np/T^3 as a function of $1/\boldsymbol{k}T$.

The slope of the straight line gives a value of 0·79 ev for the energy gap at $T=0^\circ$K (assuming that E_g is a linear function of T), with an estimated accuracy to 1%. This is in very good agreement with the most recent value obtained by Morin and Maita (1954) (0·785 at $T=0^\circ$K), from conductivity measurements in the intrinsic range. Taking $\mu_p=2000(300/T)^{2\cdot 3}$ and $\mu_n=4000(300/T)^{1\cdot 7}$ cm^2 v^{-1} sec^{-1}, the constant A was evaluated as $2\cdot 8\times 10^{32}$, the estimated accuracy being to 10%. The agreement with the value given by Morin and Maita ($3\cdot 10\times 10^{32}$) is again satisfactory.

It should be pointed out that the method described enables a direct determination of the equilibrium minority carrier concentration in the extrinsic range (equation (2)). In other measurements, such as conductivity, the majority carriers play the major role, the contribution of the minority carriers constituting a small fraction of the measured effects.

In the course of measurements it was found that in certain cases the processes of recovery (following exclusion) and decay (following injection), observed in the same section of the filament, did not behave in a completely symmetrical manner. In these cases the recovery process was not truly exponential. Such effects were encountered in a few filaments with etched surfaces, and in filaments with roughened surfaces at elevated temperatures. The nature of this unexpected behaviour is far from being understood, and further work on this problem is now being carried out.

References

Angello, S. J., and Ebert, T. E., 1954, *Phys. Rev.*, **96,** 221.
Banbury, P. C., 1953, *Proc. Phys. Soc.* B, **66,** 50.
Gibson, A. F., 1954, *Physica,* **20,** 1058.
Low, G. E., 1955, *Proc. Phys. Soc.* B, **68,** 310.
Many, A., 1954, *Proc. Phys. Soc.* B, **67,** 9.
Morin, F. J., and Maita, J. P., 1954, *Phys. Rev.*, **94,** 1525.
Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

The Infra-Red Faraday Effect due to Free Carriers in a Semiconductor

By E. W. J. MITCHELL
Physics Research Laboratories, University of Reading

Communicated by R. W. Ditchburn ; MS. received 13th June 1955

In comparing the free electron absorption in semiconductors with the Drude–Kronig theory there are two unknown parameters—a mean effective mass m^* and a damping coefficient γ. These two parameters cannot be separately determined from absorption measurements.

γ is related to the mean time τ between successive electron–phonon collisions by $\gamma = 1/2\pi\tau$. If τ is expressed in terms of the low field d.c. mobility v and the effective mass, then the latter can be calculated from the absorption results as has been done by Kahn (1955). The infra-red dispersion due to free electrons, however, is independent of τ so that the corresponding values of m^* may be determined directly.

The dispersion is small and is not measurable by the usual techniques. However, m^* could be determined from measurements of the Faraday rotation, which is linearly dependent on the dispersion. Such a measurement also has the advantage that, since the Faraday effect is entirely electronic, it does not include any contribution to the dispersion from the vibrational absorption bands.

The rotation of the plane of polarization in any optically active material is given by

$$\theta(\text{rad}) = \frac{\omega L}{\mathbf{c}} \frac{n_- - n_+}{2}$$

where ω is the circular frequency, $\mathbf{c}$ the velocity of light *in vacuo*, L the length of the specimen and $n_- - n_+$ the difference in refractive indices for the two senses of circularly polarized radiation. The Faraday rotation is related to the dispersion of plane polarized radiation, assuming that there is no interaction between the electrons, by

$$\theta = \frac{\omega L}{\mathbf{c}} \nu_{\mathrm{L}} \frac{dn}{d\nu}$$

where the Larmor precession frequency $\nu_{\mathrm{L}} = eH\mu/4\pi m^*$, e (e.m.u.) being the electronic charge, H the applied field in oersted and μ the permeability, which may be taken as unity.

The dispersion may be evaluated using the results of the free electron theory for an assembly of non-interacting electrons moving in a medium of uniform dielectric constant ϵ_0, that

$$n^2 - k^2 = \epsilon_0 - \frac{Ne^2}{\pi m^*} \frac{\mathbf{c}^2}{\nu^2 + \gamma^2}$$

and

$$nk = \frac{Ne^2}{2\pi m^*} \frac{\gamma}{\nu} \frac{\mathbf{c}^2}{\nu^2 + \gamma^2}.$$

In these equations for incident plane polarized radiation, n is the refractive index, k the extinction coefficient as defined when the complex index is $\mathcal{N} = n - ik$, and

N the concentration of free electrons. By differentiating with respect to ν and eliminating $dk/d\nu$ we have

$$n\frac{dn}{d\nu}\left(1+\frac{k^2}{n^2}\right)=\frac{N\boldsymbol{e}^2}{\pi m^*}\frac{\mathbf{c}^2}{\nu^2+\gamma^2}\left\{\frac{\nu}{\nu^2+\gamma^2}-\frac{k}{2n}\frac{\gamma}{\nu}\left(\frac{1}{\nu}+\frac{2\nu}{\nu^2+\gamma^2}\right)\right\}.$$

Fan and Becker (1951) have shown that the absorption in germanium and silicon varies as the square of the wavelength, up to $10\,\mu$. In this region, therefore, $\gamma^2 \ll \nu^2$. Further, the absorption coefficient does not rise above $500\,\mathrm{cm}^{-1}$ whereas provided it is less than $10^4\,\mathrm{cm}^{-1}$, $k^2 \ll n^2$. Then

$$\frac{k}{2n}\frac{\gamma}{\nu}\left(\frac{1}{\nu}+\frac{2\nu}{\nu^2+\gamma^2}\right)<\frac{\nu}{\nu^2+\gamma^2}$$

and the dispersion is given by

$$\frac{dn}{d\nu}=\frac{1}{n}\frac{N\boldsymbol{e}^2}{\pi m^*}\frac{\mathbf{c}^2}{\nu^3}.$$

Since $N\boldsymbol{e}=\sigma_0/v$, where σ_0 is the d.c. conductivity, and since n may be taken as $\sqrt{\epsilon_0}(=n_0)$ the Faraday rotation is

$$\theta(\mathrm{rad})=\frac{\boldsymbol{e}^2\mathbf{c}}{2\pi m^{*2}}\frac{L}{n_0}H\frac{\sigma_0}{v}\frac{1}{\nu^2}.$$

Then θ is independent of γ as compared with the absorption which is linearly dependent on γ under the same conditions. It should be possible to use a specimen of germanium 0·5 cm thick when σ_0 is $100(\Omega\,\mathrm{cm})^{-1}$, then for a field of 10^4 Oe the rotation at $5\,\mu$ is $\theta(\mathrm{deg})=0{\cdot}1/m_e^{*2}$, where m_e^* is the mean effective mass in units of free electron mass. Since $m_e^*<1$ the expected Faraday rotation is greater than 0·1°.

Of the various methods of photoelectric polarimetry it is hoped to extend into the infra-red the method, based on that due to Lyot (1948), which has been developed in this laboratory for Faraday effect measurements in the visible region.

REFERENCES

FAN, H. Y., and BECKER, M., 1951, *Semiconducting Materials*, Ed. H. K. Henisch (London: Butterworths Scientific Publications).
KAHN, A. H., 1955, *Phys. Rev.*, **97**, 1647.
LYOT, B., 1948, *C. R. Acad. Sci., Paris*, **226**, 137.

Photoeffects in Condensed Aromatic Hydrocarbons

BY D. C. NORTHROP AND O. SIMPSON
Services Electronics Research Laboratory, Baldock, Herts.

MS. received 8*th June* 1955 *and in final form* 7*th July* 1955

PHOTOCONDUCTIVITY in anthracene, and in other condensed aromatic hydrocarbons, has been observed by several workers (Inokuchi 1954, Chynoweth and Schneider 1954, and others), and has been attributed to absorption in the fundamental band of the solid. Preliminary experiments by the authors have shown that this photoconductivity can be greatly reduced by the introduction of small quantities of other hydrocarbons as impurities. We have also demonstrated a change in dielectric constant produced by light of the same wavelength. It is the purpose of this communication to outline the results obtained so far, and to suggest the importance of photoeffects in a general study of this class of solids.

Measurements were made on sandwich type specimens, prepared by evaporating the hydrocarbon on to a Pyrex flat, which had previously been coated with a transparent electrode of tin oxide. A top metal electrode was then evaporated on to the hydrocarbon, and the specimen illuminated through the transparent base. The films were polycrystalline, and microscopic examination showed the crystals to be 5 to 10 microns in diameter. Some crystals at the edges could be seen to extend through the film, and it is most probable that in all cases the specimens were only one crystal thick between the electrodes. An apparatus for measuring the electrical conductivity of such specimens has already been described (Northrop and Simpson 1954), and this has been used without modification for the photoconductivity.

In agreement with the results of previous authors, the photoconductive spectrum of anthracene was found to coincide approximately with the absorption spectrum of the crystalline solid, with maximum in the near ultra-violet. An estimate of the response to illumination with wavelengths in the neighbourhood of 4000 Å gave an increase in conductivity by a factor 10^3 due to an intensity of 5×10^{-3} w cm^{-2} in a specimen of thickness 3 microns. A phenomenon which has not been previously reported is the effect on the photoconductivity of impurities of other hydrocarbons. Specimens containing impurities of naphthacene and pentacene in concentration 1 : 1000 were found to have photoconductivity smaller than that of pure anthracene by a factor of about 100. Both these impurities also have the property of strongly quenching the blue fluorescence of anthracene (Simpson and Northrop 1954). A similar concentration of coronene, which has only a small quenching effect on the fluorescence, was found to have little effect on the photoconductivity either.

A second photoeffect was observed with the same light intensity modulated at 800 c/s. When a specimen was arranged in series with a 1 MΩ resistance and a 1·5 v battery, a signal with the periodicity of the modulation and amplitude 1–2 millivolts was generated across the resistance. This output was several orders of magnitude greater than could be accounted for by the photocurrent described above, and observation of the phase showed that it was $\pi/2$ in advance of the light signal. In this case the specimen, which has a capacity of about 800 pF, was behaving as a purely capacitive impedance of about 0·25 MΩ, and the signal was due to the modulation of the dielectric constant by light. In contrast to the photoconductive effect in anthracene, the photocapacity was not found to be markedly sensitive to impurities of other hydrocarbons. This is illustrated in figure 1 where the tail ends of the photocapacitive sensitivities of anthracene, and anthracene doped with naphthacene, are compared with the long wavelength absorption of solid anthracene.

Similar photoeffects, both photoconductive and photocapacitive, were also observed in chrysene. The long wavelength sensitivities are compared in figure 2.

These preliminary results may be summarized by the following observations:

(i) The photoconductivity of anthracene is quenched by impurities which also quench the fluorescence, but not by an impurity which leaves the fluorescence unchanged. (ii) The capacity of a layer of anthracene between electrodes may

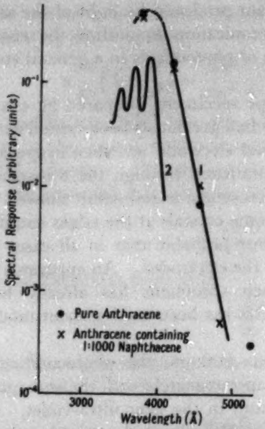


Figure 1. Photocapacity of anthracene. 1. Photocapacity. 2. Absorption of anthracene crystals (Kortüm and Finckh 1942).

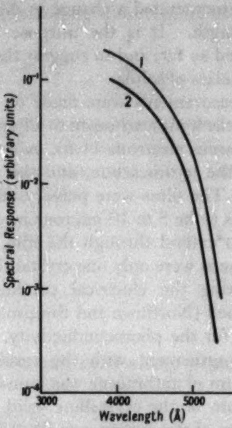


Figure 2. Photoeffects in chrysene. 1. Photoconductivity. 2. Photocapacity.

be modulated by interrupted illumination. The response time of the effect has not yet been measured, but it is certainly less than 10^{-3} second. (iii) The photocapacity is not strongly dependent on the impurity content of the specimen. (iv) The long wavelength limit of both photoconductivity and photocapacity coincides with the long wavelength limit of the fundamental absorption band of the pure solid.

On the basis of these results it is possible to infer something about the mechanism of each phenomenon. The photocapacity is probably different from that observed in inorganic phosphors (Garlick and Gibson 1946) since this was due to electrons trapped at luminescence centres, and had a relaxation time equal to the phosphorescence decay time. In the organic solids the process is not governed by impurity centres, and therefore the response time will be due to other factors.

Secondly the photoconductivity derives from absorption of light in the fundamental band of the crystal, an excitation which is normally followed by fluorescence emission. Excitation in this band does not lead to ionization of the molecules, and a non-conducting state should result. It is known that in anthracene the excitation energy may be transferred to impurity molecules, and it has been suggested that excitons are responsible for this (Bowen, Mikiewicz and Smith 1949). We make the further suggestion that the photoconductivity of pure anthracene derives from the dissociation of a small fraction of the excitons into pairs of free carriers; this might occur at the surfaces of crystals or internal cracks. Such a hypothesis would account satisfactorily for the double role of impurities in quenching fluorescence and photoconductivity, the degree of quenching being determined by the cross section of the impurity for exciton capture.

ACKNOWLEDGMENT

We wish to make acknowledgment to the Admiralty for permission to publish this paper.

REFERENCES

BOWEN, E. J., MIKIEWICZ, E., and SMITH, F. W., 1949, *Proc. Phys. Soc.* A, **62**, 26.
CHYNOWETH, A. G., and SCHNEIDER, W. G., 1954, *J. Chem. Phys.*, **22**, 1021.
GARLICK, G. F. J., and GIBSON, A. F., 1946, *Proc. Phys. Soc.*, **188**, 485.
INOKUCHI, H., 1954, *Bull. Chem. Soc. Japan*, **27**, 21.
KORTÜM, G., and FINCKH, B., 1942, *Z. Phys. Chem.*, **52**, 263.
NORTHROP, D. C., and SIMPSON, O., 1954, *Proc. Phys. Soc.* B, **67**, 892.
SIMPSON, O., and NORTHROP, D. C., 1954, *Physica*, **20**, 1122.

CORRIGENDUM

Thermal Expansion of Potassium Chlorate, by M. A. LONAPPAN (*Proc. Phys. Soc.* B, 1955, **68**, 75).

In table 5 the values for $d\beta/dt$ should be negative and not positive.

REVIEWS OF BOOKS

Méthodes de différentiation et d'intégration numériques (*applications*), par A. Ziller. Pp. 150. (Paris: Publications Scientifiques et Techniques de l'Air, N. T. No. 50, 1955.) 1500 *fr.*

This book deals primarily with interpolation and the numerical evaluation of differential coefficients and integrals of tabulated functions. For example formulae are derived which approximate an integral as a linear combination of function values at points distributed over the range of integration. These are the usual formulae based on polynomial interpolation and there is little in this part that cannot be found in a standard textbook on numerical analysis.

The main value of the book however lies in the wealth of examples given to illustrate the preliminary analytical treatment which is very often necessary before a numerical formula can be applied, as for example, in the case of an integral where the integrand or a derivative has a singularity in the range of integration (for in this case polynomial approximation breaks down). Various transformations are suggested in such cases where the analytical form of the integrand is known.

A section on numerical integration of differential equations is included. The Taylor series method is described in relation to first- and second-order differential equations. For the latter class of equation two-point boundary conditions and eigenvalue problems are discussed. R. A. BROOKER.

Dynamical Theory of Crystal Lattices, by Max Born and Kun Huang. Pp. xii + 420. (Oxford: Clarendon Press, 1954.) 50*s.*

The dynamical theory of crystal lattices shares the distinction with the kinetic theory of gases of being a comprehensive attempt to deduce the macroscopic properties of a medium from a definite microscopic model. Following on some preliminary work of Voigt, Professor Born must rank as the chief inspirer of this discipline and his contributions to it range over a period of forty years. Since its conception, Born's programme has however been somewhat modified by the realization of the great part played by the exclusion principle in all situations where the metallic bond predominates and the theory of metals has developed its own special methods. The volume under review is a comprehensive account of the lattice theory of ideal, non-conducting crystals; it supersedes the review article of Born and Mrs. Mayer, which appeared in 1933 and is in German, and gives evidence on every page, if only in manner of presentation, of the progress made since that date.

The book is divided into two parts. The first part provides in 165 pages a very readable account of such topics as atomic forces, lattice energies, lattice vibrations and the theory of specific heat both from the Einstein–Debye and Born–Kármán point of view, elasticity and stability of lattices and the dispersion theory of polar lattices.

The second part of the book is quite different in style and makes much greater demands on the reader. It is quite formal, runs to four chapters covering over 200 pages, and contains the deduction of the optical, thermodynamical and elastic properties of the infinite crystal lattice from the general potential function based on the Born–Oppenheimer adiabatic approximation.

S. ZIENAU.

(A full review of this book is published in Section A.)

Théories relativistes de la gravitation et de l'électromagnetisme, by A. LICHNEROWICZ. Pp. xii + 298. (Paris: Masson, 1955.) 2,800*fr.*

This volume in the series of mathematical works for physicists is addressed to those with a good knowledge of tensor analysis, and some acquaintance with topology. Great care has been taken to achieve the utmost rigour in the mathematics, and to avoid undesirable reliance on physical intuition. The book is divided into two parts of roughly equal length. The first part is concerned with general relativity; here the author begins by considering exactly the nature of the space–time manifold, the systems of coordinates used, and the hypotheses about the existence of derivatives of various orders. The study of the energy-tensor and the field equations leads to Cauchy's problem. (That is, given the values of the field variables and their first derivatives on a hypersurface S, to find their values inside (or outside) S.) The detailed consideration of Cauchy's problem, for the various theories, is an excellent feature of the book. In this, as in its elegance and exactness, it stands in the direct line of descent from the great French mathematical works. The interior and exterior forms of Cauchy's problem are connected by 'jump' conditions across S. The interpretation of the energy-tensor is then discussed in terms of hydrodynamics. Chapters are included on the relativistic hydrodynamics of a viscous fluid, and of a charged perfect fluid. Finally, new physical problems about the properties of the space–times in the large are considered; for example, static, everywhere regular, space–times are discussed.

The second part of the book deals with unified field theories. It opens with an excellent general introduction. The ideal of a unified field theory is traced to the combination of the electric and magnetic field strengths into a single six-vector, effected by the special theory of relativity. However, when the gravitational field receives an explanation by means of the extension to the general theory, the electromagnetic field is described in terms of concepts derived from general relativity, but by equations deduced by plausible extrapolation from the special theory. These modified Maxwell equations involve the metric, so that there is some interaction between the fields, and their laws of propagation are identical. The author concludes that such a theory, though an advance, is still only a provisional electromagnetic theory. The need is felt for a truly unified theory, in some sense. He distinguishes two senses; in the general sense a theory will be called unified if the gravitational and electromagnetic fields enter symmetrically, each coming from the same geometrical structure. In the strict sense, a theory is unified to the extent to which the exact equations represent one non-decomposable field, so that the separation into gravitational and electromagnetic field equations is possible only approximately under specially simple physical conditions.

The author divides the attempts to find a unified field theory into the two general classes of five-dimensional theories (including the so-called projective theories), and theories with an affine connection. As an example of the first class, he studies the Jordan–Thiry theory, with references to the earlier Kaluza–Klein theory. As the representative of the second class, the Einstein–Schrödinger theory is presented in a somewhat similar fashion to the presentation by Schrödinger in ' Space–Time Structure '.

The book can be earnestly recommended to specialists in general relativity who wish to consider seriously the problems raised. It is written throughout with the greatest care and precision, and printed very well; the alphabetical index, however, is wholly inadequate. C. W. KILMISTER.

Numerical Methods, by A. D. BOOTH. Pp. vii + 195. (London : Butterworths Scientific Publications, 1955.) 35*s*.

The book is mainly based on a series of lectures given to honours mathematics students at Birkbeck College, London. Its aim is to give a general introduction to the basic techniques of Numerical Analysis with special reference to those methods which are most appropriate to automatic digital computers, with the development and use of which the author has been associated for the last ten years.

After a short introductory chapter of a general nature there are three chapters on the elementary theory of the difference calculus in which the main interpolation, integration and differentiation formulae are developed mainly by operational methods. It should be noticed that the formula giving the first derivative in terms of differences is incorrect and that the incorrect formula is used later in the book. There is a special section on the gaussian integration formulae which have proved of much greater use on automatic computers than on desk machines. A general method of estimating errors in approximate integration formulae in terms of derivatives is given, but it is not stressed that in general it is of much greater value to have errors in terms of differences.

The solution of ordinary differential equations starting from given initial values is treated in one short chapter. The main method described is the Adams–Bashforth, with the Picard process or the Taylor series method for the initial steps. Although the Taylor series method is fully described no mention is made of its value as a method in its own right for solving differential equations, rather than as a starting process, though in cases where it can be used it is often extremely powerful. The Runge–Kutta process is dismissed as inferior to the Picard or the Taylor series method for starting a solution and, rather surprisingly, no mention is made of Gill's excellent adaptation of the Runge–Kutta process though it is a good example of a method which is well suited to an automatic machine while having nothing to recommend it on desk machines.

The next chapter, dealing with matrix algebra, is the least satisfactory in the book, which is unfortunate in view of the great use of such methods which has been made on automatic computers. The method of pivotal condensation for solving linear equations is described and it is stated that the virtue of taking the largest element is that it keeps rounding errors to a minimum, but that it is not well suited to automatic machines. No mention is made of two far more important advantages, that of helping to keep the elements of the reduced matrices from growing too rapidly in magnitude and of avoiding the difficulty which otherwise arises when a principal minor, but not the whole matrix, is singular or nearly singular. Both of these are of fundamental importance on an automatic machine and if the largest element in the leading column, rather than in the whole matrix, is taken as pivot this is probably the most suitable of known methods. The final stages of the Choleski method for inverting a matrix are described in such a manner as to justify the author's comment that " the complexity of the process is self-evident ". An iterative process for finding the inverse of a matrix A by means of the formula $X_{n+1}=X_n(2I-AX_n)$ is illustrated by a most misleading example which obscures the great value of the formula for improving an approximate inverse found by some other method. In the section on latent roots of matrices, unsymmetric matrices receive little attention although they are of frequent occurrence, and complex roots of real matrices are dismissed with the erroneous statement that if A has a complex root AA' will have a real root equal to the square of the modulus of that of A.

By contrast the chapter on partial differential equations, which is mainly devoted to the quasi-linear equation of the second order, is well written and within the limits imposed by the space available is reasonably complete. The classification of second order equations into elliptic, hyperbolic and parabolic types by reduction to canonical form is described and there is an exposition of the main methods of treating the three types of equation, in which matrix methods, relaxation techniques and Monte-Carlo methods are fully described. The method of characteristics for solving hyperbolic equations receives less attention than it deserves however, and it is implied that it is only when the characteristics can be found analytically that the method is useful. This is rather surprising in view of the fact that the method has been extensively used on problems in compressible flow in which the characteristic systems are often very complex. Although the use of higher differences for estimating errors in results is described in several contexts there is no reference to the extensive work of L. Fox in this field.

The remaining chapters are devoted to short accounts of non-linear algebraic equations, the use of approximating functions, Fourier synthesis and integral equations. Each of these gives a concise account of the principal techniques involved, perhaps the only omission worthy of note being Bairstow's process for finding the complex zeros of real polynomials.

There is at present no other book which covers the whole range of topics treated in this volume, and workers in the field of numerical analysis will find a useful introduction to their subject together with fairly extensive references to the literature. Unfortunately, the work is marred by a surprising number of errors of a careless nature and this, together with the price, precludes an unconditional recommendation. J. H. WILKINSON.

Millimicrosecond Pulse Techniques, by I. A. D. LEWIS and F. H. WELLS. Pp. xiv + 310. (London : Pergamon Press, 1954.) 40*s*.

This book is one of a series of monographs on Electronics and Waves. As the authors explain in the preface, the subject of the book is formed by pulse techniques characterized by time-intervals of 10^{-6} to 10^{-10} second. Continuous wave techniques are not included.

There are eight chapters, which in the reviewer's opinion can be divided into three groups. The first, and shortest, consists of the first chapter—the Theoretical Introduction. Here there are some definitions and a breath-taking, though adequate, introduction (five pages) to Laplace transforms.

The second group comprises the chapters on transmission lines and transformers. The former contains a very interesting section on helical lines, as well as treatments of uniform lines and lumped delay lines; there is a good summary of use of lines in pulse-shaping and also a discussion of discontinuities and terminations. The chapter on transformers is almost wholly concerned with various forms of line transformers to which thirty-five pages are given. The lumped pulse transformers are dismissed after two pages, the reason given being that they are adequately described elsewhere. The references given are mostly to various papers, including a Telecommunications Research Establishment monograph which may be difficult to obtain. In the reviewer's opinion, a more detailed treatment of lumped pulse transformers would have been well worth while.

Both the above chapters are of a physical character and deal with the design of lines and transformers. They are developed from the standpoint of electromagnetic theory, with considerable use of mathematical techniques. The remaining chapters, which form the third group, are mainly concerned with circuit techniques ; the chapter headings are : Pulse Generators, Amplifiers, Cathode Ray Oscilloscopes, Applications to Nuclear Physics, and lastly Miscellaneous Applications.

The chapter on amplifiers contains a well-developed section on distributed amplifiers with discussion of gain stability, noise, and a summary of the performance of existing types. There is also a section on the salient properties and applications of secondary emission valves. This section is particularly welcome, as practically all otherwise available information is contained in original papers. Unfortunately confusion occurs on a rather vital point. On page 142 a method of stabilizing mutual conductance is shown, with the comment that " the circuit . . . is subsidiary to, and must not interfere with, the normal functioning as a high-speed amplifier ". The reader is not told how to make it not interfere, while the circuit diagram shows condensers of unstated value connected straight across the four a.c. electrodes.

The chapter on oscillographs contains a very lucid treatment of the effects of transit-time and of ringing in the deflector plate circuits. Further, various special types of cathode-ray tubes are described and there is a detailed discussion of spot size. Lastly representative time-base circuits are given. This is perhaps the best written chapter in the book.

Next come the two chapters on applications, that on applications to nuclear physics being significantly the longest of all chapters. After a summary of the various kinds of nuclear radiation detectors, there follows a description of discriminators, scaling circuits, and above all coincidence circuits.

The volume is very well produced, there are plenty of diagrams and a long list of references. The mistakes are few and of a trivial nature ; thus on page 138 there is a reference to a non-existing paragraph 5.2.4(b), which should presumably read 5.2.1.3.

The book is well worth reading for all concerned with the millimicrosecond field. As there is a lot of general discussion in the text, the book should remain up to date for quite a long time, although the sections on applications will necessarily age.

A. FOLKIERSKI.

Electronics, by T. B. Brown. Pp. xi + 545. (New York: John Wiley, 1954.) 60*s*.

This sizeable volume covers a wide field beginning with diodes and ending with ultra-high-frequency electronics. There are fourteen chapters in all. The first seven deal with diodes, triodes, tetrodes and pentodes, voltage and power amplifiers, feedback circuits, radio frequency amplifiers and oscillators. Three chapters are given to electron emission and the description and application of gas valves. There are further chapters on modulation and on ultra-high-frequency electronics, and two chapters on selected topics under the titles of Special Functions and Electronic Instruments.

The book is well written and well produced ; it makes interesting reading. The reviewer however has found it irritating to have the text repeatedly interrupted by sections called Laboratory Experiments and Demonstration Experiments. These cannot be omitted in reading (especially the latter), as cross-references occur in the text.

A number of topics is treated more extensively than is usual in books of this scope. This is true, for instance, of the chapters on gas valves. The chapter on negative feedback is very thorough, and the concept of output impedance is very rightly stressed. Graphical analysis is used extensively, and is particularly enlightening in the unusually comprehensive treatment of oscillators. Another valuable topic covered is the effect of a.c. voltages on several grids in one valve.

Some useful circuits, surprisingly, do not appear ; amongst these the most prominent are the virtual earth amplifier, the Miller integrator, and the diode-coupled scale-of-two (flip-flop) circuit. Instead a rather unusual scale-of-two circuit is described. In chapter 10 rotating vector diagrams are introduced. Unfortunately the direction of measurement of voltages is not indicated on the circuit diagrams. While this fault is very common amongst authors, it is very regrettable. It makes the direction of a voltage vector arbitrary, as any such vector could equally well be drawn in the opposite phase. Incidentally, the author uses different arrow-tips for voltage and current vectors ; it is particularly sad that in the first diagram in which this convention is illustrated a mistake occurs, and a voltage vector masquerades with a ‘ current ’ arrow-tip.

The book is addressed to students. It will probably appeal only to the serious ones, who are not too mathematically minded. For the stress is on concepts and the presentation is graphical, in so far as possible. The exposition of the subject is very carefully thought out, and the volume should form a valuable basis for those who may be planning to deliver courses of lectures on electronics.

A. FOLKIERSKI.

CONTENTS OF SECTION A

PAGE

Mr. H. Trefall. On the Barometer Effect on the Hard Component of the Cosmic Radiation . . . 953

Dr. R. F. Barrow, Mr. D. V. Glaser and Dr. P. B. Zeeman. Rotational Analysis of Bands of the $A^3\Pi_{0^+}, B^3\Pi_1 - X^1\Sigma^+$ Systems of Indium Monofluoride . . . 962

Dr. J. N. Murrell. The Electronic Spectrum of Aromatic Molecules: VI—The Mesomeric Effect . . . 969

Mr. J. Hubbard. The Dielectric Theory of Electronic Interactions in Solids . 976

Dr. S. L. Altmann. The Self-Consistent Field for Zr^{4+} . . . 987

Mr. D. M. Brink. Inelastic Scattering of Neutrons by Deformed Nuclei . 994

Prof. W. E. Burcham, Dr. J. L. Symonds and Mr. J. D. Young. The Production of ^{11}C from Carbon by Protons of Energy between 200 and 950 mev . . 1001

Dr. J. M. Calvert, Dr. A. A. Jaffe, Mr. A. E. Litherland and Mr. E. E. Maslin. Neutron Groups from Deuteron Stripping by ^{23}Na, ^{27}Al and ^{31}P . . 1008

Dr. J. M. Calvert, Dr. A. A. Jaffe and Mr. E. E. Maslin. An Investigation of the Reaction $^{19}F(d, n)^{20}Ne$ using a new Fast-Neutron Spectrometer. . 1017

Dr. B. Donovan. The Hall Effect in Metals at High Frequencies 1026

Mr. M. J. Laubitz. The Beta Decay of the Ground State of ^{27}Al 1033

Dr. G. R. Satchler. Gamma Radiation following the Surface Scattering of Nucleons . . . 1037

Dr. G. R. Satchler. Polarization in Nuclear Reactions 1041

Mr. J. L. Lloyd and Dr. A. W. Wolfendale. The Scattering of μ-Mesons in Lead and Iron . . . 1045

Dr. G. C. Fletcher. Calculation of Magnetostriction Constants for Nickel . 1066

Research Notes:

Dr. J. W. Gardner. Transport of Energy by Knock-on Electrons . . 1072

Dr. W. D. Allen and Dr. A. T. G. Ferguson. The n–p Cross Section in the Range 60–550 kev . . . 1077

Mr. D. A. Hancock and Mr. F. Verdaguer. The Absolute Determination of Resonant Energies for Radiative Capture of Protons by Sodium in the Energy Range below 530 kev . . . 1080

Mr. R. Batchelor and Mr. G. C. Morrison. The Angular Distribution of $^7Li(p, n)^7Be^*$ Neutrons for Proton Energies between 2·5 and 2·9 mev 1081

Mr. M. A. Perks. The Quadrupole Moments of Odd *A* Nuclei . . . 1083

Letters to the Editor:

Prof. C. B. A. McCusker and Mr. B. G. Wilson. A Search for Secondary Maxima in Cosmic Ray Transition Curves: I . . . 1086

Mr. P. D. McCormack, Prof. C. B. A. McCusker and Mr. B. G. Wilson. A Search for Secondary Maxima in Cosmic Ray Transition Curves: II 1088

Dr. J. M. Baker and Dr. B. Bleaney. The Nuclear Magnetic Moment of Holmium 165 . . . 1090

Corrigenda: A. D. Buckingham and J. A. Pople . . . 1091
A. D. Buckingham . . . 1091

Reviews of Books . . . 1092

Contents of Section B . . . 1096

The Lifetime of Added Carriers in InSb

By I. M. MACKINTOSH† and J. W. ALLEN‡
† Department of Physics, University of Nottingham
‡ Ericsson Telephones Ltd., Beeston, Nottingham

Communicated by L. F. Bates; MS. received 18*th March* 1955
and in amended form 29*th August* 1955

Abstract. The Roosbroeck–Shockley theory of radiative recombination of electrons and holes in semiconductors is extended to the case of degenerate semiconductors and applied to InSb. The experimental optical absorption curves of various n-type and p-type samples are analysed and results given for the lifetime of added carriers for this process alone as a function of the position of the Fermi level. An extrapolation indicates a decay time for a small disturbance in carrier concentration in intrinsic material of 0·75 μsec. Comparison with experimental results indicates that this process may be predominant in InSb.

§ 1. INTRODUCTION

RECENT experimental observations have indicated that the lifetime of added carriers in the intermetallic semiconductor InSb is considerably less than in the other adamantine semiconductors so far investigated. It was therefore thought to be of interest to obtain a theoretical value for the photon-radiative recombination rate, since this may be expected to be an important process in InSb.

The mechanisms by which the energy of recombination of an electron–hole pair may be released are: (*a*) direct recombination with photon emission, (*b*) direct recombination with phonon emission, (*c*) recombination via one or more intermediate levels with photon, phonon, or photon–phonon emission, (*d*) Auger effect.

Theoretical considerations indicate that process (*b*) is unimportant, and since there is no information at present available concerning intermediate energy levels in InSb no estimate can be made of the importance of process (*c*). Pincherle (unpublished work) has recently investigated the importance of the Auger effect as a cause of recombination in semiconductors, and from the expression obtained by him the recombination rate in InSb due to this process may be shown to be negligible. For example, the lifetime due to the Auger effect in intrinsic InSb is calculated to be 3×10^{-3} sec.

Recombination by the emission of a single photon may, however, be estimated theoretically using the theory of Roosbroeck and Shockley (1954) which applies the principle of detailed balance to the production of electron–hole pairs by photon absorption.

§ 2. THEORY

At thermal equilibrium the principle of detailed balance gives the result that the rate of radiative recombination for a small range of frequencies $d\nu$ at a frequency ν is equal to the rate of generation of electron–hole pairs by thermal radiation of interval $d\nu$ and frequency ν. If $\rho(\nu)\,d\nu$ is the density in the crystal of photons in the frequency range $d\nu$, and $P(\nu)$ is the probability per unit time that a photon of frequency ν will be absorbed by excitation of an electron from the valence band to the conduction band, then the rate of generation per unit volume and unit frequency interval is $P(\nu)\rho(\nu)$. Hence the total rate of generation R per unit volume is

$$R=\int_0^\infty P(\nu)\rho(\nu)\,d\nu. \qquad \ldots\ldots(1)$$

Roosbroeck and Shockley have shown that this may be expressed in the form

$$R=\frac{32\pi^2}{c^3}\left(\frac{\boldsymbol{k}T}{h}\right)^4\int_0^\infty n^3K\frac{u^3}{e^u-1}\,du \qquad \ldots\ldots(2)$$

where $u\equiv h\nu/\boldsymbol{k}T$, n is the refractive index, and K is the absorption index given by

$$K=\alpha\lambda/4\pi n \qquad \ldots\ldots(3)$$

where α is the absorption coefficient. For InSb, the refractive index may be assumed constant and equal to 3·75 over the wavelength range considered (Avery *et al.* 1954).

For the value at 300°K, equation (2) reduces to

$$R=1{\cdot}78\times10^{22}\int_0^\infty n^3K\frac{u^3}{e^u-1}\,du\ \mathrm{cm^{-3}\,sec^{-1}}. \qquad \ldots\ldots(4)$$

The lower limit of integration is the value of u which corresponds to the long-wavelength absorption limit, about 7μ for pure InSb. Longer wavelength photons cannot produce electron–hole pairs and are absorbed by other mechanisms. By the principle of detailed balance R is also the total rate of photon-radiative recombination of electrons and holes.

The derivation of the relation between the lifetime and recombination rate for a non-degenerate semiconductor has been given by Roosbroeck (1953). His result may be expressed in the form

$$\frac{1}{\tau_p}=R\left(\frac{1}{n}+\frac{1}{p}\right) \qquad \ldots\ldots(5)$$

where τ_p is the decay time of a small number δp of holes in an n-type semiconductor, and n and p are the electron and hole concentrations respectively. It follows that the decay time in material of intrinsic conductivity will be

$$\tau_i=n_i/2R \qquad \ldots\ldots(6)$$

where n_i is the concentration of intrinsic electrons.

InSb, however, is degenerate to quite low impurity concentrations and to low temperatures (Burstein 1954), and can be very strongly degenerate at room temperature. In the case of n-type degeneracy, only those electrons lying close to the Fermi level E_F will normally be able to take part in the recombination processes, and it is not justifiable to assume simply that n is large enough, relative to p, to be ignored in (5). A simple analysis has shown that account may be taken of this fact in the following way.

For the case of complete degeneracy,

$$n=8\pi\{2m_e(E_F-E_c)\}^{3/2}/3h^3 \qquad \ldots\ldots(7)$$

where E_c is the energy of the bottom of the conduction band and m_e is the free electron mass. Hence the effective concentration of electrons n_F near the Fermi level can be obtained by the usual approximation,

$$n_F\sim n\boldsymbol{k}T/(E_F-E_c). \qquad \ldots\ldots(8)$$

The decay time for degenerate material is then obtained by substituting n_F for n in equation (5).

Calculations then show that for the n-type samples considered in this paper, $n_F \gg p$, and hence $1/n_F$ is negligible compared with $1/p$. As an example, for a very impure sample of InSb for which $E_F - E_c = 0{\cdot}30$ ev, expression (8) gives $n_F = 3 \times 10^{17}$ cm^{-3}, whereas $p = 1{\cdot}9 \times 10^{10}$ cm^{-3}.

Thus for strongly degenerate n- or p-type InSb, the decay times for minority holes or electrons reduce to

$$\tau_p = p/R, \quad \tau_n = n/R. \qquad \ldots\ldots(9)$$

§ 3. Results

The designation of the specimens the authors of the absorption curves used and other relevant data are shown in table 1. The position of the Fermi level for specimens 1 and 2 was obtained from the absorption data by extrapolating the experimental curves to zero absorption. It was then assumed that this point corresponds to transitions to the lowest available energy levels, which, in accordance with the work of Burstein (1954), may be assumed to be about $4kT$ below E_F. This gives values of $(E_F - E_v)$ where E_v is the energy of the top of the valence band, from which n_F and p can be calculated. For specimens 3 and 4 the position of the Fermi level was calculated from the published carrier concentrations.

Table 1. Theoretical Estimates of the Lifetime for Various InSb Specimens

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	Avery *et al.* (1954)	*n*	3·3	0·48	$3{\cdot}3 \times 10^{17}$	$1{\cdot}9 \times 10^{10}$	$2{\cdot}15 \times 10^{20}$	$8{\cdot}8 \times 10^{-11}$
2	Breckenridge *et al.* (1954)	*n*	5·5	0·33	$2{\cdot}3 \times 10^{17}$	$6{\cdot}2 \times 10^{12}$	$1{\cdot}29 \times 10^{21}$	$4{\cdot}8 \times 10^{-9}$
3	Avery *et al.* (1954)	*n*	7·2	0·23	$1{\cdot}3 \times 10^{17}$	$3{\cdot}0 \times 10^{14}$	$1{\cdot}20 \times 10^{22}$	$2{\cdot}5 \times 10^{-8}$
4	Breckenridge *et al.* (1954)	*p*	7·2	0·08	$2{\cdot}7 \times 10^{15}$	$1{\cdot}0 \times 10^{17}$	$0{\cdot}98 \times 10^{22}$	$2{\cdot}4 \times 10^{-7}$

(1) Specimen; (2) authors; (3) conductivity type; (4) absorption edge (microns); (5) $E_F - E_v$ (ev); (6) n or n_F (cm^{-3}); (7) p (cm^{-3}); (8) R (cm^{-3} sec^{-1}); (9) τ (sec).

Figure 1 shows the form of the dependence of $U \equiv u^3/(e^u - 1)$, n^3K, and Un^3K on u for specimen 1, the most impure InSb sample. It is only for this specimen that an absorption curve is available over a sufficiently large wavelength range to give the flattening-out part of the n^3K curve at higher values of u. The n^3K curves for the other specimens are essentially similar except for a variation in the position of the absorption edge, but an extrapolation has had to be made at the higher values of u. The inaccuracies thereby introduced are likely to be small, since the fall in the factor U is already predominant in this region.

Since it is only absorption due to transitions from the valence to the conduction band which are of interest, these curves must be corrected by subtracting the effect of free carrier absorption. It was found that the latter could be represented by an empirical expression of the form

$$K = A\lambda^3 + B\lambda \qquad \ldots\ldots(10)$$

where A and B are constants. After applying this correction it was found that the absorption edge, plotted as $\log n^3K$ against u, was a straight line, and it is of interest to note that this implies that in the region of the absorption edge

$$\alpha \propto \nu e^{\gamma\nu} \qquad \ldots\ldots(11)$$

where γ is a constant and where the refractive index is assumed constant.

The results of the numerical integration of the graphs of

$$P(\nu)\rho(\nu) = 1{\cdot}78 \times 10^{22} Un^3K$$

of the various specimens are shown in column 8 of table 1, and the results for the lifetime of added minority carriers as calculated from equation (5) are shown in column 9.

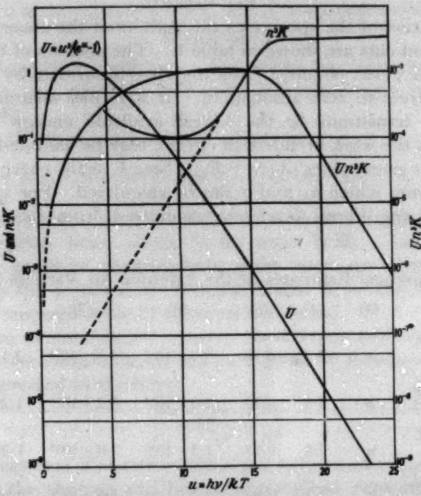


Figure 1. Dependence of n^3K, $U \equiv u^3/(e^u - 1)$ and Un^3K on $u \equiv h\nu/kT$.

The values for R of the pure n-type and pure p-type specimens are almost equal, as would be expected, since these are not strongly degenerate, and hence for intrinsic material, the value $R = 1{\cdot}1 \times 10^{22}\,\mathrm{cm}^{-3}\,\mathrm{sec}^{-1}$ may be assumed. Using the relation

$$n_i^2 = N_c N_v \exp[-(E_c - E_v)/\boldsymbol{k}T] \qquad \ldots\ldots(12)$$

and the values $m_n = 0{\cdot}03 m_e$, $m_p = 0{\cdot}2 m_e$, $E_c - E_v = 0{\cdot}18$ ev (Burstein 1954), it is found that $n_i = 1{\cdot}67 \times 10^{16}\,\mathrm{cm}^{-3}$ at 300°K. Hence the value for the radiative lifetime of intrinsic InSb at 300°K is calculated to be

$$\tau_i = 0{\cdot}75\,\mu\mathrm{sec}. \qquad \ldots\ldots(13)$$

This may be regarded as the upper limit for the recombination lifetime in this material.

The theoretical dependence of the photon-radiative lifetime and the factors $(1/n) + (1/p)$ and R on the position of the Fermi level are shown in figure 2. As expected the lifetime has a maximum value for intrinsic material.

The accuracy of the values of the radiative lifetime represented in figure 2 is dependent on the following factors: (*a*) absorption data which are possibly inaccurate, (*b*) the assumptions necessary to obtain a full curve of n^3K against u, as the available absorption data cover an insufficient wavelength range, (*c*) uncertainties in the value of p due to possible errors in the position of the Fermi level and the value of the effective mass. We have considered the possible range of the n^3K curves due to factors (*a*) and (*b*) and come to the conclusion that the value of the numerical integral will be in error by less than a factor of 2.

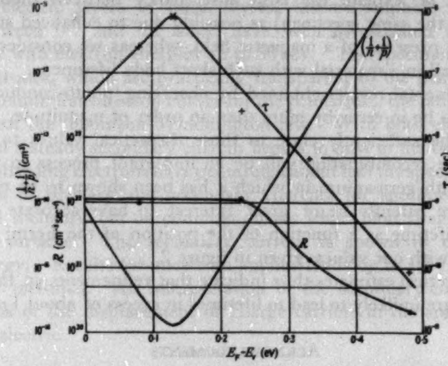


Figure 2. Dependence of $(1/n)+(1/p)$, R, and τ on the position of the Fermi level.

For intrinsic material the position of the Fermi level is known quite accurately, whereas in the other cases it is not possible to estimate accurately the uncertainties in p, since the experimental data is incomplete. It is therefore considered that the result for the intrinsic lifetime is accurate to within a factor 2, with a larger uncertainty in the other values.

§ 4. DISCUSSION OF RESULTS

There is little experimental evidence so far available on the lifetime in InSb, and the accuracy of the few values obtained is doubtful due to the difficulty of measuring such short lifetimes with present methods and techniques. Published values at 300°K are shown in table 2.

Table 2. Experimental Values of the Lifetime for Various InSb Specimens

Author	Method	Description of specimen	Lifetime (sec)
Moss (1954)	Photovoltage and photoconductivity	n-type and polycrystalline	$2{\cdot}5\times10^{-8}$
Kurnick *et al.* (1954)	PEM	Intrinsic	$\sim10^{-10}$
Kurnick *et al.* (1954)	Photo-conductivity	Intrinsic	$\sim10^{-8}$

The result obtained by Moss by measurement of photo-voltaic and photo-conductive effects was obtained with polycrystalline material in which the diffusion length was of the same order of magnitude as the crystallite dimensions. Thus it is open to doubt whether his result represents a true body lifetime.

The values obtained by Kurnick *et al.* (1954) are open to criticism on the following grounds. There is some uncertainty in the interpretation of the photo-electro-magnetic (PEM) effect in InSb since the value of the effective mobility given by Aigrain and Bulliard (1953) will be widely different from that given by Moss *et al.* (1953) when the ratio of the carrier mobilities is large, as is the case in this material. This may therefore invalidate their first value. The criticisms made of the value obtained by Moss may also apply to their second value since it is not stated that the material used was single crystal.

Kurnick *et al.* explain the large discrepancy between their two results (obtained with the same specimen) as possibly due to enhanced surface recombination in the presence of a magnetic field, whereas we consider this effect is likely to be small in a material with such short body lifetimes.

The experimental results obtained by observing photo-conductive changes are not likely to be in error by more than an order of magnitude, however, and thus seem to be sufficiently close to these theoretical values to suggest that photon-radiative recombination will be an important process in this material, as compared with germanium in which it has been shown to be negligible. It would therefore patently be of great interest to have accurate experimental values of the lifetime as a function of the position of the Fermi level, and to compare these with our values, given in figure 2.

These theoretical estimates thus indicate that refinements in the techniques of preparation are unlikely to lead to lifetimes in excess of about 1 μsec in InSb.

ACKNOWLEDGMENTS

One of the authors (I. M. M.) wishes to express his thanks to Professor L. F. Bates for the facilities of his laboratory and for guidance and supervision, to Dr. K. W. H. Stevens for helpful discussions, and to the Chief Scientist, Admiralty, for permission to publish this work. The other (J. W. A.) is indebted to Dr. J. H. Mitchell, Controller of Research, Ericsson Telephones Ltd., for permission to participate in this work.

REFERENCES

AIGRAIN, P., and BULLIARD, H., 1953, *C.R. Acad. Sci., Paris,* **236,** 595.

AVERY, D. G., GOODWIN, D. W., LAWSON, W. D., and MOSS, T. S., 1954, *Proc. Phys. Soc.* B, **67,** 761.

BRECKENRIDGE, R. G., BLUNT, R. F., HOSLER, W. R., FREDERIKSE, H. P. R., BECKER, J. H., and OSHINSKY, W., 1954, *Phys. Rev.*, **96,** 571.

BURSTEIN, E., 1954, *Phys. Rev.*, **93,** 632.

KURNICK, S. W., STRAUSS, A. J., and ZITTER, R. N., 1954, *Phys. Rev.*, **94,** 1791.

MOSS, T. S., 1954, *Proc. Phys. Soc.* B, **67,** 775.

MOSS, T. S., PINCHERLE, L., and WOODWARD, A. M., 1953, *Proc. Phys. Soc.* B, **66,** 743.

VAN ROOSBROECK, W., 1953, *Phys. Rev.*, **91,** 282.

VAN ROOSBROECK, W., and SHOCKLEY, W., 1954, *Phys. Rev.*, **94,** 1558.

Note added in proof. Since the acceptance of this paper experimental results agreeing closely with our calculated intrinsic lifetimes have been published by D. G. Avery and D. P. Jenkins (*J. Electronics*, 1955, **1,** 145).

Electron Bombardment Effects in Thin Dielectric Layers

By W. E. SPEAR
Physics Department, University College of Leicester

MS. received 19*th May* 1955 *and in final form* 18*th July* 1955

Abstract. Dielectric layers of mica, Pyrex, arsenic sulphide and antimony sulphide, between 0·7 and 4 μ thick, have been investigated under electron bombardment with electron energies up to 50 kev. The specimens carry evaporated Al electrodes, which are practically transparent to the primary electrons. The results include transmission curves for the primaries; the range values are in agreement with the Thomson–Whiddington law. With both electrodes at the same potential a steady secondary current begins to flow to the bottom electrode when the bombarding electrons have penetrated about half the specimen. Experimental values of primary and secondary currents for specimens of different thickness and dielectric fall on common curves when plotted against a suitable independent variable. The secondary current is shown to depend on the absorbed energy. Results for mica with applied fields up to 10^6 vcm^{-1} across the specimen are included. An interpretation of the secondary current curves is given in terms of the displacement of charge carriers in the space-charge field within the dielectric.

---

§ 1. Introduction

In recent years a number of papers have been published dealing with the steady currents which flow in thin dielectric layers under bombardment by high-speed electrons. Pensak (1949) has investigated electron-bombardment conductivity effects in thin films of amorphous silica; Ansbacher and Ehrenberg (1949, 1951) have described work on arsenic sulphide and aluminium oxide films (the 1951 paper will be referred to as AE below).

The experiments showed that under the action of an applied field across the specimen the secondary currents may rise to a large multiple of the incident current. The work of AE and Pensak is mainly concerned with the investigation of the secondary phenomena in the region of intermediate and high current gains. Little attention has been given to processes connected essentially with the primary electrons in their passage through the dielectric, which nevertheless are important in the interpretation of electron-bombardment conductivity effects. In this paper both primary and secondary effects are studied and in particular it is attempted to link the two. So far the experiments have been confined to the region of zero or low current gain. It has been found essential to distinguish more carefully between primary electrons and secondary charge carriers than has been done in the above-mentioned experiments, which were carried out on specimens deposited on a thick base. The preparation of unbacked specimens of arsenic sulphide is briefly described in § 3. In the absence of an applied field the charge carriers produced by the primary beam drift in the space-charge field, so that the measured secondary current curves throw some light on the space-charge distribution in a thin dielectric under electron bombardment.

§ 2. EXPERIMENTAL ARRANGEMENT

The bombarding electron beam enters through a hole in the hemispherical electrode H (figure 1). The beam is focused magnetically to a circular spot of 0·26 cm diameter, which just passes through the hole in the aluminium top electrode T′ of the specimen holder. Electrons transmitted by the specimen S pass through the aperture B′ (diameter 0·7 cm) into the collector C. The current to each of the electrodes can be measured by plugging a galvanometer into the jack J incorporated in each lead. It was found most convenient to employ two instruments, one in jack A, measuring the total incident electron current i_p, the other to read the current to the electrode under consideration at accurately constant i_p. It will be noticed that i_p includes only electrons absorbed and transmitted by S, and not those which are re-diffused and leave the top surface of the specimen with a considerable fraction of their incident energy. Most of these electrons are collected by H and only a negligible fraction by T′. The latter is normally kept at sufficiently negative potential to suppress true secondary emission from the top surface T of the specimen and to return slow tertiaries to H. The same applies to the aperture B′ which faces the bottom electrode B of the specimen. In some of the experiments a field E was produced across the specimen by applying a potential V_B, or bias, to B. In that case $V_{B'}$ was adjusted so as to keep B′ about 120 v negative with respect to B. The collector C consists of an aluminium tube of 9 cm length and 3 cm diameter. It collects most of the transmitted current, except that at higher accelerating voltages when most of the primary electrons are transmitted, B′ receives about 8% of the incident current consisting of electrons with energies greater than 500 ev. It appears most likely that these are transmitted electrons which have been re-diffused from the walls of C. In the following experiments a suitable correction has been applied to the value of i_c.

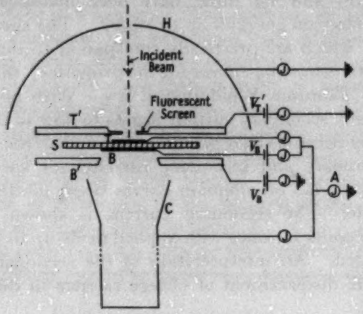


Figure 1. Experimental arrangement and measuring circuit. S, specimen; T, B, evaporated top and bottom electrodes of specimen; T′, B′, specimen holder; C, collector; J, jacks.

§ 3. PREPARATION AND MOUNTING OF SPECIMENS

The dielectrics investigated were Pyrex, mica, arsenic sulphide and antimony sulphide. The specimen is sandwiched between evaporated aluminium electrodes B and T as shown in figure 2. The thickness of the electrodes is of the order of

3×10^{-6} cm (8 μg cm^{-2}). As the electron range energies of the specimens are about 10 kev, the error in the determination of the transmitted current due to absorption at the electrodes is negligible.

The specimen is mounted between two mica washers M, each of which has a central hole of the same size as that in aperture B′. The contacts to the evaporated electrodes are made by thin strips of aluminium foil F, pressed by the washers to the ends of T and B. The mounted specimen is screwed to electrode T′ and can be reversed.

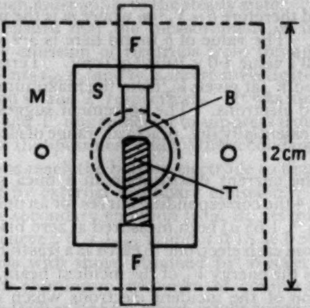


Figure 2. Mounted specimen. S, specimen; T, B, top and bottom electrodes; F, contact foils; M, mica frame.

Pyrex films of about 1 μ thickness were obtained from blown bulbs. It was not found possible to cleave a mica specimen of less than about 4 μ thickness. A number of arsenic sulphide specimens were prepared by a method similar to that used for thin unbacked gold films of accurately known thickness (Spear 1955). It was found best to evaporate the complete specimen including the aluminium electrodes on thin electrolytically polished copper foil. The specimen is then mounted in the mica frame and a drop of nitric acid applied to the copper. The backing dissolves within a few seconds; the aluminium is not attacked. As electrode B covers the full aperture of the mica frame, no acid comes into contact with the dielectric. It is rather difficult to obtain specimens which show little or no measurable leakage current when tested with an applied field of the order of 10^5 v cm^{-1}. A few of the arsenic sulphide specimens were satisfactory in this respect and it was possible to obtain complete transmission characteristics. Other experiments on arsenic and antimony sulphide films were made with composite specimens which consisted of an aluminium base (beaten or evaporated foil) of about 0·8 μ thickness, on to which the dielectric and the top electrode were evaporated.

The evaporated films were made from B.D.H. Analar reagents. According to AE the arsenic sulphide films are amorphous whilst antimony sulphide films have a microcrystalline structure.

§ 4. RESULTS AT ZERO BIAS

The performance of the apparatus was tested using an evaporated aluminium foil of 0·316 mg cm^{-2} in the specimen holder. If the fraction i_c/i_p, where i_c is the transmitted electron current collected by C, is plotted against the energy of the incident electrons, a curve very similar to the i_c curves in figures 3 and 4 is obtained. Extrapolation to $i_c/i_p=0$ of the linear portion leads to a value $V_p=V_0$. A good deal of experimental evidence (Whiddington 1914, Terrill 1923, Schonland 1925) suggests that the loss of velocity of electrons in their passage through thin metallic foils is given to a good approximation by the Thomson–Whiddington law. This relation can be stated in terms of the incident electron energy and the energy V_x which the electrons possess at a depth x within the foil.

Thus
$$V_p^2-V_x^2=b\rho x. \qquad \ldots\ldots(1)$$
At $V_p=V_0$, $V_x=0$ and therefore $b=V_0^2/m$ where m is the mass per unit area of the foil and ρ the density. The value of b found here is $3{\cdot}9\times10^{11}\,\mathrm{v^2\,g^{-1}\,cm^2}$ in good agreement with Terrill's value $4{\cdot}0\times10^{11}\,\mathrm{v^2\,g^{-1}\,cm^2}$. Terrill did not use the range method but determined V_x at given V_p from the maximum of the energy distribution of the transmitted electrons. The agreement suggests that the value of V_0 found here represents essentially the ' practical ' range of electrons of most probable energy.

Figure 3 shows the current distribution in a mica specimen ($1{\cdot}3\,\mathrm{mg\,cm^{-2}}$, $4{\cdot}3\,\mu$ thick), and figure 4 the corresponding curves for an unbacked arsenic sulphide specimen ($0{\cdot}56\,\mathrm{mg\,cm^{-2}}$, $1{\cdot}65\,\mu$) both measured at zero bias. The ordinate gives the electron current from each electrode to earth as a fraction of the incident current i_p, the abscissa denotes the energy V_p of the incident beam, in kev; $i_B/i_p=1-i_c/i_p$ represents that fraction of the incident electrons which end their paths within the specimen. The i_c curve is accurately reproducible; the initial ordinates of the i_B curve of a new specimen are normally 5–10% higher than those obtained after about half an hour's bombardment. After this the curves are reproducible.

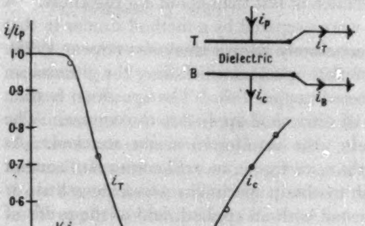


Figure 3. Current distribution in a mica specimen, $1{\cdot}3\,\mathrm{mg\,cm^{-2}}$. Constant incident current, $i_p=3{\cdot}56\times10^{-8}$ A.

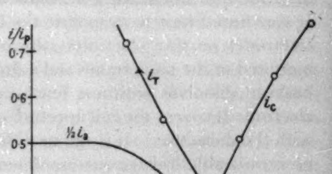


Figure 4. Current distribution in an arsenic sulphide specimen, $0{\cdot}56\,\mathrm{mg\,cm^{-2}}$. Constant incident current, $i_p=3{\cdot}56\times10^{-8}$ A.

The fraction i_B/i_p was found to be independent of the current density if the latter is varied over a range of $1:10^3$ at a bombarding energy corresponding to the maximum of the i_B curve.

It is interesting to note in connection with figures 3 and 4 that an electron current flows to the bottom electrode before any of the incident electrons have sufficient energy to penetrate the dielectric; i_B shows a sharp cut-off corresponding to an energy V_0' of the incident electrons. The i_B and i_T curves are perfectly symmetrical about the $\frac{1}{2}i_a$ line; this is to be expected from the fundamental condition $i_a = i_B + i_T$ which must apply to the steady state.

It is important to ensure that the measurement of internal effects is not influenced by secondary emission of electrons from top and bottom electrodes. For this reason the currents i_B, $i_{B'}$ and i_T, $i_{T'}$ were investigated for various positive and negative potentials on B′ and T′ respectively. It was found that (*a*) a potential of − 120 v applied to B′ and T′ was sufficient to suppress true secondaries from top and bottom electrodes of the specimen, and (*b*) the secondary emission starts when the first primary electrons reach the bottom electrode so that the initial part of the i_B curve is not affected by external secondary emission. In figure 3 the curve marked i_s represents the secondary emission from B; its initial portion coincides with the 'toe' of the i_c curve; $(i_s/i_p)_{\max}$ is about 0·18 at $1{\cdot}55\,V_0$. It is of interest to compare these values with those obtained by Wecker (1941, tabulated by Bruining (1954)) for an aluminium specimen of comparable thickness (3·5 μ). Wecker finds $(i_s/i_p)_{\max} = 0{\cdot}09$ at $1{\cdot}65\,V_0$, which might suggest that in our case some secondaries produced near the mica interface are able to reach the surface through the aluminium film.

Dielectric	m (mg cm^{-2})	V_0 (kev)	V_0' (kev)	b (v^2 g^{-1} cm^2)	$V_0'^2/m$ (v^2 g^{-1} cm^2)	x'/d	Remarks
Mica	1·30	26·6	20·6	$5{\cdot}45 \times 10^{11}$	$3{\cdot}28 \times 10^{11}$	0·60	
Mica	1·10	24·6	19·2	$5{\cdot}5 \times 10^{11}$	$3{\cdot}35 \times 10^{11}$	0·61	
Pyrex	0·815	22·5	15·9	$6{\cdot}2 \times 10^{11}$	$3{\cdot}1 \times 10^{11}$	0·50	
Pyrex	0·15	9·65	6·75	$6{\cdot}2 \times 10^{11}$	$3{\cdot}1 \times 10^{11}$	0·50	
Pyrex	0·18	10·5	7·56	$6{\cdot}1 \times 10^{11}$	$3{\cdot}2 \times 10^{11}$	0·52	
As_2S_3	0·56	16·4	12·05	$4{\cdot}8 \times 10^{11}$	$2{\cdot}58 \times 10^{11}$	0·53	unbacked
As_2S_3	0·29	11·8	8·35	$4{\cdot}8 \times 10^{11}$	$2{\cdot}4 \times 10^{11}$	0·50	on evaporated foil
As_2S_3	0·31	12·1	8·8	$4{\cdot}7 \times 10^{11}$	$2{\cdot}5 \times 10^{11}$	0·53	on beaten foil
Sb_2S_3	0·525	16·0	11·7	$4{\cdot}9 \times 10^{11}$	$2{\cdot}6 \times 10^{11}$	0·53	on beaten foil

The current distribution curves of figures 3 and 4 are typical of all the specimens investigated. In the table the values of V_0, V_0', and b for further specimens can be found. The V_0 and V_0' columns show that in every case the i_B current begins to flow before complete penetration has taken place, which suggests that this effect is a general property of dielectric films under electron bombardment independent of structure or thickness.

The value of b is constant for each substance within the accuracy of the experiment which shows that an expression of the form of equation (1) holds for dielectrics, though with a higher value of b than for metal foils. At a bombarding voltage $V_p = V_0'$, the primary electrons have, according to the Thomson–Whiddington law, penetrated to a depth $x' = V_0'^2/b\rho = (V_0'/V_0)^2 d$. The experimental x'/d values shown in the table indicate therefore that a steady current of slow electrons flows in a bombarded dielectric film if the primary beam has penetrated to about half the depth of the specimen. In view of the energy distribution of the transmitted electrons the above statement refers essentially to electrons having the most probable energy. An estimate based on the extent of the 'toe' of the i_c/i_p curve in figure 3 indicates that for the more energetic electrons the x'/d values may be up to 15% higher than those tabulated.

It is well known (cf. Bothe 1933) that the true range of the electrons (i.e. the total integrated path length R) is about twice the practical range, so it is not improbable that an electron with energy $V_p = V_0'$ might reach the bottom electrode. In this case one would expect a rapid increase of i_c for all $V_p > V_0'$; but in all experiments it is found that $i_c = 0$ up to very nearly $V_p = V_0$, so that the onset of the i_B current cannot be connected with the condition $R = d$.

The above results can be compared with figure 6 of the paper AE. There, the square of the critical voltage at which the electron-bombardment conductivity current in arsenic sulphide starts to flow is shown to be proportional to the film thickness. The graph correlates the results for about twenty specimens between 1 μ and 4 μ thick, and leads to a gradient of 0·78 which corresponds to a range constant of $2{\cdot}4 \times 10^{11}\,\mathrm{v^2\,g^{-1}\,cm^2}$ ($\rho = 3{\cdot}4\,\mathrm{g\,cm^{-3}}$). This agrees well with the values of $V_0'^2/m$ in the table and shows fairly conclusively that the interpretation of the critical voltage given in AE as the true range of the bombarding beam is incorrect. The table includes results for composite specimens which were obtained as follows: at first the electron range $(V_0)_{\mathrm{Al}}$ of the aluminium backing was determined before evaporation. Then i_c and i_B curves were measured on the complete specimen with (a) the thin top electrode and (b) the base towards the incident beam. The combined range $(V_0)_c$ agreed exactly in both cases. Equation (1) leads to $V_0 = [(V_0)^2_c - (V_0)^2_{\mathrm{Al}}]^{1/2}$; V_0' can be found directly from (a).

It is of interest to try the correlation of complete current distribution curves for different dielectrics and thicknesses. The constancy of $(V_p/V_0)^2$ for the beginning of the i_B current (see x'/d column in table) suggests the use of $(V_p/V_0)^2$ rather than V_p as the independent variable. Figure 5 shows that now all the $\frac{1}{2}i_a/i_p$ values fall on a common curve. It has not been found possible to correlate in the same way the i_B curves which have been measured at constant i_p. If, however, at each V_p the value of i_p is so chosen that W_p, the incident energy per second, remains constant, then some degree of correlation is obtained as indicated by the i_B/W_p curve in figure 5. Values for specimens of the same dielectric but corresponding to different thicknesses fall reasonably well on a common curve. These results will be discussed in § 6.

§ 5. RESULTS WITH APPLIED FIELD

Unfortunately the unbacked arsenic sulphide specimens develop a permanent 'leakage' when fields greater than about $10^5\,\mathrm{v\,cm^{-1}}$ are applied between T and B. It has therefore not been possible to link up the results at $E = 0$ in the case of arsenic sulphide with those obtained by AE with applied fields of the order of $10^6\,\mathrm{v\,cm^{-1}}$. However, mica and Pyrex specimens were more suitable in this respect. Figure 6 shows the current distribution in a mica specimen at fields up to $10^6\,\mathrm{v\,cm^{-1}}$. E is positive in the positive x direction (i.e. T to B). The curves are very similar in appearance to those obtained by AE but the current gains are considerably lower; there is no significant shift of V_0' with E. Figure 6 shows that with a negative applied field $i_T = 0$ at some value $V_p = (V_p)_1$ and becomes negative for all $V_p > (V_p)_1$. The change of sign of i_T is of interest because it seems to mark roughly the beginning

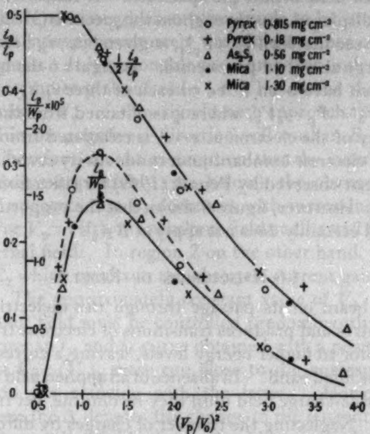


Figure 5. Correlation of i_a/i_p and i_B/W_p values for specimens of different thickness and dielectric. Broken line: i_B/W_p values for Pyrex specimens; full line: i_B/W_p values for mica and arsenic sulphide specimens.

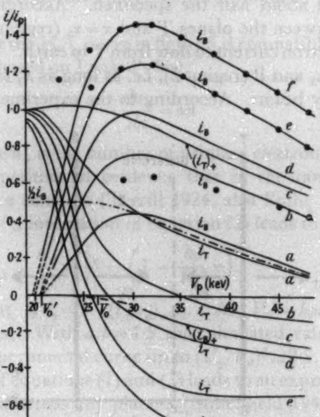


Figure 6. Current distribution in mica specimen, 1·3 mg cm⁻², at different applied fields, E in v cm⁻¹; *a*, $E=0$; *b*, $E=-2{\cdot}8\times10^5$; *c*, $E=+4{\cdot}7\times10^5$; *d*, $E=-4{\cdot}7\times10^5$; *e*, $E=-8{\cdot}1\times10^5$; *f*, $E=-10{\cdot}5\times10^5$. The marked points denote the total absorbed energy per second fitted to the i_B/i_p curves at $V_p=48$ kev.

of a uni-directional field within the specimen. It follows that under these conditions all charge displacements throughout the specimen are linked with B (and T). It is interesting to see that for $V_p > (V_p)_1$, at given bias, i_B becomes proportional to W_a, the total energy absorbed per second. In figure 6 the marked points denote values of W_a which have been fitted to each of three i_B/i_p curves at $V_p = 48$ kev. W_a is given by $W_a = W_p - i_c V_d$, where i_c is obtained from the transmission curve and V_d, the energy of the electron at $x = d$, is calculated from equation (1). The proportionality of electron-bombardment conductivity currents and total absorbed energy has also been observed by Pensak (1949) for silica films at applied fields of about 10^6 v cm^{-1}. However, figure 6 shows that the proportionality breaks down at lower fields and certainly does not apply at $E = 0$.

§ 6. Discussion of Results

The primary beam on its passage through the dielectric releases electrons from trapping centres and produces transitions of electrons from the filled energy band of the insulator to higher energy levels, leaving a corresponding number of positive holes in the filled band. In absence of an applied field these charge carriers will drift in the space-charge field until they recombine, get trapped, or reach one of the electrodes. Neglecting the transfer of charges by diffusion, it will depend on the field produced by the space charge whether a particular region in the specimen will contribute a current to the top or to the bottom electrode.

The experimental evidence suggests that at some level $x = x_1$ the space-charge field changes sign (figure 7). Otherwise it would be difficult to explain the observed i_B curves which in all cases begin to rise sharply as soon as the primary beam has penetrated about half the specimen. According to figure 7 charge carriers produced between the planes T and $x = x_1$ (region 1) will drift in such a way as to cause an electron current to flow from T to earth. $i_B = 0$ if no secondaries are created between x_1 and B (region 2), i.e. as long as $x_1 \geqslant x_p$, the depth of penetration of the primary beam. According to the experimental data x_1 should be approximately $\frac{1}{2}d$ at $V_p = V_0'$. As x_p exceeds x_1 charge displacements will occur in region 2, which produce an external electron current flowing away from B. i_B will rise with increasing penetration provided the plane x_1 does not move appreciably towards B. For $V_p > V_0$, i_B passes through a maximum which

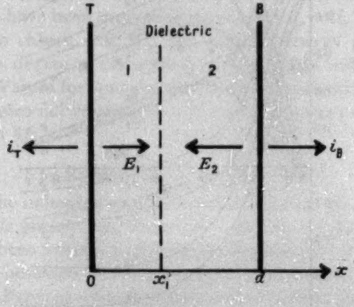


Figure 7.

represents the optimum combination of carrier production and space-charge field in region 2. At larger V_p the energy absorption tends to be uniform throughout the specimen, so that now approximately the same number of secondaries are produced in both regions. It is reasonable to assume that under these conditions the field approaches symmetry about $x/d = \frac{1}{2}$. This means that i_B should tend towards i_T with increasing V_p, i.e. they should both approach the $\frac{1}{2}i_a$ line.

An external field in the negative x-direction will weaken E_1 and cause a shift of x_1 towards T. Therefore, on the assumption that the external field does not modify the space charge appreciably (see below), the change of sign of i_T at $V_p = (V_p)_1$ (figure 6) suggests that E_1 decreases with increasing V_p in the range $V_0' < V_p < (V_p)_1$. For $V_p > (V_p)_1$ the resultant field at all x will begin to act in the direction of the external field. In region 2 on the other hand, the same external field will reinforce E_2 which results in the observed current gain shown by the i_B curves in figure 6. The approximately constant value of V_0' indicates that for $V_p < (V_p)_1$ the resultant field must be controlled predominantly by space charge.

Figure 6 also shows an i_B and i_T curve obtained with a positive field (denoted by $(i_B)_+$, $(i_T)_+$). At $V_p > (V_p)_1$ these run close to the corresponding $(i_T)_-$ and $(i_B)_-$ curves, but do not coincide with the latter. However, in all cases $\Delta(i_B)_- = \Delta(i_T)_+$, where the Δ denotes the change in the current from that at zero bias. This means that applied fields of considerable magnitude in either direction appear to be additive to the original space-charge field at $E = 0$, although with the increased displacement of carriers one might expect a change in charge density. One possible explanation is that at $E = 0$ the contribution of the secondary charge carriers to the resultant charge density is small, so that the space-charge field is predominantly controlled by the charge density produced by absorbed primary electrons.

It can be shown that the $\frac{1}{2}i_a/i_p$ curve in figure 5 is compatible with the calculation of electron absorption based on Lenard's law,

$$-\frac{dn_x}{dx} = \alpha n \qquad \ldots\ldots(2)$$

(cf. Bothe 1933), where n_x is the number of primary electrons crossing plane x per second. There is experimental evidence that in the range of V_p used here $\alpha = a\rho/V_x^2$, where a is a constant (Terrill 1924, also Bothe 1933). Substituting from equation (1) for V_x, integration of equation (2) leads to

$$\frac{n_x}{n_p} = \frac{i_x}{i_p} = \left(1 - \frac{b\rho}{V_p^2}x\right)^{a/b}. \qquad \ldots\ldots(3)$$

The absorbed current $i_a/i_p = 1 - (i_x/i_p)_{x=d}$. As $V_0^2 = b\rho d$, $\frac{1}{2}i_a/i_p = f(V_0^2/V_p^2)$ only, for all specimens. With $a/b = 1{\cdot}5$ the calculated values are in reasonable agreement with the experimental curve up to $(V_p/V_0)^2 \simeq 2{\cdot}5$.

The combination of equations (1) and (3) leads to an expression for the primary energy flow per second across any plane x (Stinchfield 1946, quoted by Garlick 1949, Koller and Alden 1951),

$$W_x = W_p\left(1 - \frac{b\rho}{V_p^2}x\right)^{a/b+1/2}$$

where W_p is the incident energy per second. Therefore the energy absorbed between $x \simeq 0{\cdot}5d$ and $x = d$ becomes $W_p f(V_0^2/V_p^2)$. The i_B/W_p curve in figure 5 suggests that at the higher $(V_p/V_0)^2$ values i_B becomes proportional to the energy absorbed per second in region 2. Obviously this does not apply at lower $(V_p/V_0)^2$ where the i_B values are largely dependent on the space-charge field. It may be of significance in this connection that the arsenic sulphide and mica specimens, whose normalized i_B values fall reasonably well on a common curve at all $(V_p/V_0)^2$, have approximately the same dielectric constant, whereas that of the Pyrex specimens is about 40% lower.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. W. Ehrenberg of Birkbeck College, University of London, for his interest in this work and for several helpful discussions.

REFERENCES

ANSBACHER, F., and EHRENBERG, W., 1949, *Nature, Lond.*, **164,** 144 ; 1951, *Proc. Phys. Soc.* A, **64,** 362.

BOTHE, W., 1933, *Handb. Phys.* XXII, 2 (2nd Edn).

BRUINING, H., 1954, *Physics and Applications of Secondary Electron Emission* (London : Pergamon Press).

GARLICK, G. F. J., 1949, *Luminescent Materials* (Oxford : Clarendon Press), p. 194.

KOLLER, L. R., and ALDEN, E. D., 1951, *Phys. Rev.*, **83,** 684.

PENSAK, L., 1949, *Phys. Rev.*, **75,** 472 ; 1950, *Ibid.*, **79,** 171.

SCHONLAND, B. F. J., 1925, *Proc. Roy. Soc.* A, **108,** 187.

SPEAR, W. E., 1955, *J. Sci. Instrum.*, **32,** 34.

TERRILL, H. M., 1923, *Phys. Rev.*, **22,** 101 ; 1924, *Ibid.*, **24,** 616.

WECKER, F., 1941, *Ann. Phys., Lpz.*, **40,** 405.

WHIDDINGTON, R., 1914, *Proc. Roy. Soc.* A, **89,** 554.

The Binding Energy of Nitrogen in a Dislocation

BY W. R. THOMAS† AND G. M. LEAK
The British Iron and Steel Research Association, Metallurgy (General) Division, Sheffield

MS. received 30th June 1955 *and in amended form 29th August* 1955

Abstract. The solubility of interstitially dissolved nitrogen has been measured both in annealed and in cold-worked alpha iron by the internal friction technique in more detail than in the original work of Dijkstra. The dissolved nitrogen in annealed material, in equilibrium with a precipitated nitride, has a heat of solution of 0·26 ev per nitrogen atom. This agrees reasonably well with other published values. A new approach has been introduced into the treatment of the results for the solubility of nitrogen in cold-worked iron. Equilibrium is considered between nitrogen atoms in free solution and nitrogen atoms bound in dislocation sites. Since the number of dislocation sites is fixed, account is taken of the proportion of these occupied at different temperatures. The value for the binding energy of nitrogen atoms in dislocation sites is estimated as lying between 0·75 and 0·8 ev per nitrogen atom.

§ 1. INTRODUCTION

THE basis of the dislocation theory of the yield point in metals and the strain ageing phenomena exhibited, for example, in body-centred cubic alpha iron is the fact that solute atoms can relieve the stresses around dislocations (Cottrell 1948, 1953 a). The condition for this is that the solute atoms can migrate to suitable sites in the distorted lattice in the immediate vicinity of the dislocations (Cottrell and Bilby 1949). The driving force for this migration arises from the energy of interaction between the stress fields of the dislocation and the solute atom. In the interesting case of a supersaturated solid solution the dislocation atmosphere is in competition with the precipitate particles for the solute atoms (Cottrell and Leak 1952). The relative binding energies of the solute in dislocation atmospheres and in precipitated particles will influence the final site for the solute atoms.

Considering specifically the case of carbon and nitrogen in alpha iron, these solutes can interact with both screw (Nabarro 1948) and edge components (Cottrell and Bilby 1949) of dislocations. These atoms occupy possible interstitial sites at the centres of {001} faces or the mid-points of ⟨001⟩ edges of the body-centred cubic cells. This introduces tetragonality by an extension along the ⟨001⟩ edge. The direction of the tetragonal axis can be changed by a small movement of the interstitial atom through a distance of only $a/2$ where a is the lattice parameter. Consequently interstitial solutes can relieve shear stresses in body-centred cubic lattices and so can interact with screw dislocations. This jump process also forms the basis of Snoek's (1941) treatment of the internal friction peak in alpha iron at low frequencies.

---

† Now at Department of Mines and Technical Surveys, Atomic Energy Project, Chalk River, Ontario, Canada.

The volume change introduced into the lattice by a solute atom can relieve the stress around an edge dislocation. For a solute atom at a point with polar coordinates r, α from a positive edge dislocation then the binding energy V between the dislocation and the solute atom is given by

$$V = -\frac{A \sin \alpha}{r} \qquad \ldots\ldots(1)$$

where A depends upon the elastic constants of the material, the volume change introduced by the solute, and the dislocation strength. Where this volume change is a dilation of the lattice, then V will be a maximum for $\alpha = 3\pi/2$ and r approximately equal to the interatomic distance. A can be taken to be about 3×10^{-20} dyn cm^2; thus the maximum value of V will be about 1 ev. The assumptions made for this calculation imply that this value will undoubtedly be too high. An estimation of the binding energy can be deduced from yield point measurements. The upper yield corresponds to the stress required to pull a dislocation away from its atmosphere of bound solute atoms. Cottrell (1953 b) showed that an estimate could be made of the temperature T_0 below which sites of strongest binding $V_{\max}$ were nearly all filled by solute atoms:

$$T_0 = \frac{V_{\max}}{\boldsymbol{k} \ln (1/C_0)}. \qquad \ldots\ldots(2)$$

Here C_0 is the average concentration of solute atoms in the matrix, $\boldsymbol{k}$ is Boltzmann's constant. Mott (1952) pointed out that above a temperature of about 700°K the yield point is not observed in iron. Assuming that this is a result of the small yield point due to the absence of an atmosphere of solute atoms, T_0 must be about 700°K. For C_0 equal to 10^{-4} wt.% then $V_{\max}$ is equal to about 0·5 ev. This value of C_0 is about the minimum carbon content for which a yield point has been observed at temperatures up to 400–450°C, i.e. approximately 700°K.

An experimental value of $V_{\max}$ can also be obtained directly from measurements of carbon or nitrogen solubility in cold-worked alpha iron. Dijkstra (1949) measured this solubility for nitrogen together with the solubility in unworked iron, in equilibrium with precipitated iron nitride. He used the internal friction technique developed by Snoek. The solubility in unworked iron was 4–5 times greater than that in cold-worked iron, for temperatures of 300°C and 400°C. The inference is that the solution pressure of nitrogen in the precipitate is greater than in dislocations, and hence that $V_{\max} > V(\mathrm{Fe_4N})$. Dijkstra showed that $V(\mathrm{Fe_4N})$ was about 0·35 ev. From the difference in solubilities Cottrell (1953 a) estimated that $V_{\max}$ should be about 0·1 ev greater than $V(\mathrm{Fe_4N})$, and hence $V_{\max} \sim 0{\cdot}45$ ev.

Cottrell's estimate was based on only two measurements by Dijkstra. The purpose of the present note is to report a more detailed investigation of this problem by the same technique as that used by Dijkstra. The concentration C of a solute in equilibrium with a precipitated phase is given by the equation

$$\ln C = -\frac{\Delta H}{RT} + \text{constant} \qquad \ldots\ldots(3)$$

where ΔH is the heat of solution of the precipitated phase. For the specific case of nitrogen in iron in equilibrium with iron nitride ΔH is equal to the heat of solution per gramme atom of nitrogen, i.e. the energy difference between a nitrogen atom in the precipitate and one in the matrix. ΔH can immediately be estimated from a plot of $\ln C$ against $1/T$.

Superficially, it would appear that an estimate could be made of the binding energy V by using an equation, analogous to equation (3), of the form

$$\ln C_W = -\frac{V}{RT} + \text{constant} \qquad \ldots\ldots(4)$$

where V is defined as the difference in energy between a nitrogen atom in a dislocation and one in the bulk matrix. The concentration C_W represents the amount of nitrogen in solution in equilibrium with dislocation atmospheres in cold-worked iron. However, equation (4) does not take into account the equilibrium that must exist between nitrogen atoms in the matrix, nitrogen atoms bound in dislocation sites and unoccupied sites for nitrogen atoms in dislocations.

§ 2. EXPERIMENTAL DETAILS

The experimental aspect of internal friction measurement and the interpretation of the results have been discussed in detail elsewhere by, for example, Polder (1945), Dijkstra (1947) and Zener (1948). Only a summary of the essential items will be given here. The logarithmic decrement of the free torsional oscillations of an iron wire is measured as a function of temperature and the presence of carbon or nitrogen in solution gives rise to a peak in the decrement–temperature curve. The height of this peak is proportional to the amount of impurity in interstitial solid solution in the lattice.

Decarburized and denitrided wires were loaded with nitrogen using ammonia–hydrogen mixtures previously described by Thomas and Leak (1955). The high purity iron (B.I.S.R.A. Reference AHN) had the following analysis: C 0·0038 wt. %, Si 0·001 wt. %, Mn 0·005 wt. %, Cr 0·0013 wt. %, Al 0·0027 wt. %, N_2 0·0014 wt. %. The nitrogen content of the wires used in the present experiments was 0·012 wt. % and their carbon content was less than the minimum which could be detected by chemical or internal friction analysis, i.e. <0·0005 wt. %.

Wires were quenched into cold water (20°c) from the temperature of maximum solubility of nitrogen in alpha iron (580°c). Measurements were made for two series of specimens (*a*) in unworked iron, (*b*) after 20% extension in a tensile testing machine and for both series after ageing at various temperatures between 100°c and 270°c. When further ageing made no difference to the height of the relaxation peak then the value of the decrement corresponded to the equilibrium solubility of nitrogen at the temperature of ageing. Ageing was carried out in constant temperature oil baths, with a temperature fluctuation of less than ± 1°c.

A roughly constant frequency of oscillation was used throughout these series of measurements, about 1·5 to 2 c/s, which gave an internal friction peak at about 30°c. For this reason the proportionality constant between peak height and nitrogen content remained the same throughout the measurements and in the diagrams; 1 unit of logarithmic decrement is equivalent to 0·00039 wt. % of nitrogen. This factor was determined by comparing chemical analyses with internal friction measurements under conditions where all the nitrogen was in free solid solution.

§ 3. RESULTS

Ageing curves are shown schematically in figure 1. The measurements of equilibrium solubility are given in the table. The solubility values in equilibrium with a precipitated nitride in annealed material are given for purposes of comparison. In accordance with equations (3) and (4) these results are plotted in figure 2 in the form log (concentration) against $1/T$. The slopes of these curves give values for $\Delta H = 6 \pm 2$ kcal per gramme atom of nitrogen $= 0{\cdot}26 \pm 0{\cdot}09$ ev per nitrogen atom, and $V = 7{\cdot}5 \pm 2$ kcal per gramme atom of nitrogen $= 0{\cdot}35 \pm 0{\cdot}09$ ev per nitrogen atom.

Temp. of ageing (°c)	After quench ageing Wt. % × 10³	After straining and ageing Wt. % × 10³	Temp. of ageing (°c)	After quench ageing Wt. % × 10³	After straining and ageing Wt. % × 10³
90	0·92	—	142	1·9	1·1
100	—	0·42	142	—	1·3
100	—	0·55	165	3·85	1·8
108	0·94	0·73	182	4·5	2·1
108	—	0·6	201	7·9	3·4
112	1·75	0·55	201	—	4·3
120	1·4	—	230	8·9	5·0
130	—	1·05	250	9·7	9·6

In order to compare these results with those obtained by Dijkstra the effect of altering the strain was measured. No significant difference in solubility was observed in varying the strain from 20% to 45% extension, nor did the method of straining, i.e. drawing or extension, affect the results.

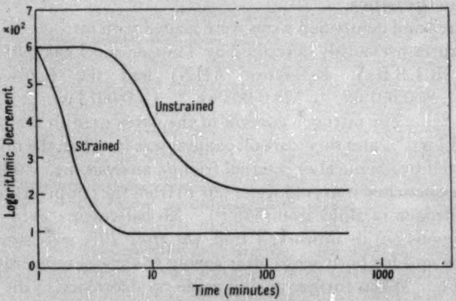


Figure 1. Representative ageing curves for 200°c.

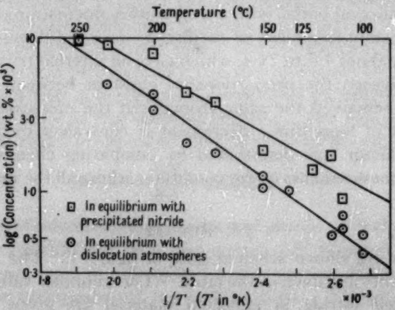


Figure 2. Solubility of nitrogen in pure iron plotted as log (concentration) against $1/T$.

§ 4. Discussion

The value for ΔH of 0·25 ev compares with Dijkstra's value of 0·35 ev; Dijsktra's results, however, had a considerable spread which would include 0·25 ev. This value, furthermore, is in close agreement with the estimation of Borelius and his co-workers (1951, 1954). The estimate of V, from the present results based on equation (4), is low compared with Cottrell's calculation from Dijkstra's results and also with Mott's calculation.

The derivation of equation (3) follows from the reaction taking place

$$4\,[\mathrm{Fe}]+[\mathrm{N}]\ \text{dissolved}=[\mathrm{Fe_4N}]. \qquad \ldots\ldots(5)$$

This gives an equilibrium constant K defined by

$$K=\frac{[\mathrm{Fe}]^4\,[\mathrm{N}]}{[\mathrm{Fe_4N}]} \qquad \ldots\ldots(6)$$

and hence ΔH given by

$$\Delta H=-R\frac{d(\ln K)}{d\,(1/T)}. \qquad \ldots\ldots(7)$$

Under the conditions of the experiment it is probable that the precipitate is $\mathrm{Fe_{16}N_2}$ (Jack 1955, private communication). This does not affect the analysis which applies equally well to either form of the nitride.

Since the proportion of nitrogen atoms to iron atoms in the iron nitride is fixed, then K is proportional to the dissolved nitrogen concentration and equation (3) follows from equation (7).

The corresponding derivation, for equilibrium between dissolved nitrogen and nitrogen associated with dislocation atmospheres can be based on the reaction

$$[\mathrm{N}]+[\mathrm{D}]=[\mathrm{ND}] \qquad \ldots\ldots(8)$$

where [D] corresponds to the free dislocation sites which could be occupied by nitrogen atoms if the reaction were to go completely to the right in equation (8). [ND] corresponds to the occupied sites in the dislocations. Hence the equilibrium constant is given by

$$K=\frac{[\mathrm{N}]\,[\mathrm{D}]}{[\mathrm{ND}]} \qquad \ldots\ldots(9)$$

and the binding energy of nitrogen in a dislocation is given by

$$V=-R\frac{d(\ln K)}{d\,(1/T)}. \qquad \ldots\ldots(10)$$

Thus equation (4) is only an approximation based on the assumption that the concentration of available dislocation sites for nitrogen atoms does not vary with temperature and that K is proportional to the concentration of dissolved nitrogen. This assumption is incorrect because it is known that, although the concentration of dislocations does not vary between room temperature and about 400°c, the existence of a pronounced yield point at room temperature and the absence of one at 400°c would indicate that the number of dislocation sites for nitrogen atoms varies markedly in this temperature range. A more precise value for V can be obtained from equations (9) and (10) in the following way. The values for [N] (in wt. % nitrogen) can be measured at various temperatures (see table); hence the value for [ND] (in wt. % nitrogen) can be obtained by difference. It remains to determine a value for [D]. Let A be the maximum possible value for [ND]; this is about 0·03 wt. % nitrogen assuming a dislocation density of 10^{12} lines per cm^2 for heavily worked material and an atmosphere of two nitrogen atoms per iron atom plane threaded by each dislocation (Cottrell 1953 *b*). Then the value of [ND] at any temperature T given by

$$[\mathrm{ND}]_T = A - [\mathrm{N}]_T \qquad \ldots\ldots(11)$$

Since the change in measured solubility $\Delta[\mathrm{N}]$ due to a variation in temperature must be equal to the change in the free dislocation sites $\Delta[\mathrm{D}]$

$$\Delta[\mathrm{N}] = \Delta[\mathrm{D}]. \qquad \ldots\ldots(12)$$

Hence $[\mathrm{D}]_T = [\mathrm{N}]_T - \{[\mathrm{N}]_{T_0} - [\mathrm{D}]_{T_0}\}$

$$= [\mathrm{N}]_T - B \qquad \ldots\ldots(13)$$

where T_0 is some arbitrary temperature. Thus

$$K(T) = \frac{[\mathrm{N}]_T\{[\mathrm{N}]_T - B\}}{A - [\mathrm{N}]_T}. \qquad \ldots\ldots(14)$$

The problem now is to adjust B so that $\ln K$ varies linearly with $1/T$. For the present results the arbitrary T_0 was taken as 105°c where $[\mathrm{N}]_{105}$ is equal to $0{\cdot}05 \times 10^{-2}$ wt. % nitrogen and $[\mathrm{ND}]_{105}$ is thus equal to $1{\cdot}15 \times 10^{-2}$ wt. % nitrogen. In figure 3 the relationship between $\log K$ and $1/T$ has been plotted assigning values for $[\mathrm{D}]_{105}$ as 0·01%, 0·1%, 1% and 10% of A; linearity is observed for $[\mathrm{D}]_{105}$ equal to 1% of A. The slopes of the straight portions of all the curves give V equal to about 0·8 ± 0·05 ev per nitrogen atom.

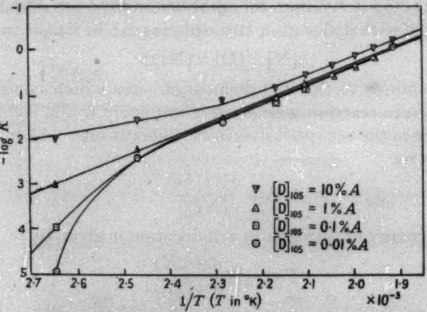


Figure 3. Equilibrium constant plotted against $1/T$.

Possible errors in this determination may arise from two sources: (i) the value taken for the total nitrogen content of the material and (ii) the estimate of the maximum available nitrogen sites in dislocations. These will affect the terms B and A. Possible variations have been plotted in figure 4. It can be seen that the slopes of these lines vary only very slightly with large changes in atmosphere concentration and nitrogen content of the specimens.

It can thus be concluded that the value for V at the dislocation core should be about 0·8 ± 0·05 ev per nitrogen atom. This figure is considerably different from the value estimated from more elementary considerations (equation (4)). The method presented here takes into account the equilibria which must exist between the possible sites for nitrogen atoms.

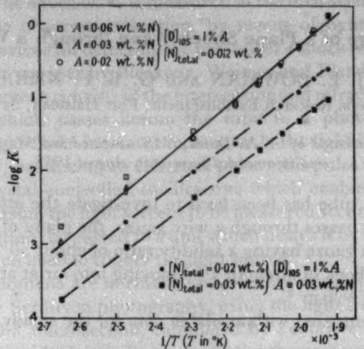


Figure 4. Equilibrium constant plotted against $1/T$.

ACKNOWLEDGMENTS

The authors are grateful to Professor A. H. Cottrell for his encouragement and interest during this work. They appreciate discussions with Dr. F. D. Richardson and, in particular, his helpful suggestions for the theoretical treatment of the results. They also wish to make acknowledgment to Mr. B. Fitton who carried out much of the experimental work.

REFERENCES

BORELIUS, G., 1951, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **197**, 477.
BORELIUS, G., and ASTRÖM, H. U., 1954, *Acta Met.*, **2**, 547.
COTTRELL, A. H., 1948, *Strength of Solids* (London : Physical Society), p. 30; 1953 a, *Progress in Metal Physics* (London : Butterworths Scientific Publications), Vol. **4**; 1953 b, *Dislocations and Plastic Flow in Crystals* (Oxford : Clarendon Press), pp. 99–150.
COTTRELL, A. H., and BILBY, B. A., 1949, *Proc. Phys. Soc.* A, **62**, 49.
COTTRELL, A. H., and LEAK, G. M., 1952, *J. Iron St. Inst.*, **172**, 301.
DIJKSTRA, L. J., 1947, *Philips Res. Rep.*, **2**, 357; 1949, *Trans. Amer. Inst. Min.* (*Metall*) *Engrs*, **194** (*J. Metals*, **1**), 252.
MOTT, N. F., 1952, *Imperfections in Nearly Perfect Crystals* (New York : John Wiley).
NABARRO, F. R. N., 1948, *Strength of Solids* (London : Physical Society), p. 38.
POLDER, D., 1945, *Philips Res. Rep.*, **1**, 5.
SNOEK, J. L., 1941, *Physica*, **8**, 734.
THOMAS, W. R., and LEAK, G. M., 1955, *J. Iron St. Inst.*, **180**, 155.
ZENER, C., 1948, *Elasticity and Anelasticity of Metals* (Chicago : University Press).

The Passage of a Plane Shock Wave through a Wire Gauze

By J. E. BOWMAN and G. B. F. NIBLETT
Atomic Weapons Research Establishment, Fort Halstead, Sevenoaks, Kent

Communicated by A. H. Davis; *MS. received* 3*rd March* 1955
and in amended form 16*th August* 1955

Abstract. A shock tube has been used to investigate the attenuation of a plane shock wave when it passes through a wire gauze, the plane of which is parallel to the shock front. A gauze having a solidity ratio of about 50% reduced the overpressure behind a shock of strength 1·65 moving into air at atmospheric pressure from 9·4 lb in^{-2} to 7·9 lb in^{-2}.

This result is compared with available data for the steady flow of air through screens.

§ 1. INTRODUCTION

The shock tube is an instrument which is becoming of increasing importance and utility for the study of transient gas flow problems. It consists, in essence, of a rigid tube divided into two sections by a gas-tight diaphragm; rupture of the diaphragm when a pressure difference exists across it initiates the gas flow to be studied.

The shock wave processes occurring in such a tube were first studied by Payman and Shepherd (1941) in this country. They conducted a series of experiments in which copper-foil diaphragms were used to sustain high pressures in the compression chamber, and the processes consequent upon rupture of the diaphragm were observed by means of a Schlieren photographic system and a wave speed camera.

During the last war shock tubes were extensively developed in America for shock wave studies. They were first used to produce shock waves for the dynamic calibration of piezoelectric pressure gauges for use in blast measurements. In 1945 Smith extended their use in his classical study of the regular and Mach reflection of plane shocks in air.

Since 1945 shock tubes have been used to an increasing extent to tackle a large range of shock wave problems. This development has been due, largely, to Bleakney and his co-workers at Princeton; they have developed a school of shock tube studies and have employed a Mach–Zehnder interferometer† to turn the shock tube into an accurate quantitative tool.

§ 2. THE A.W.R.E. SHOCK TUBE

The Atomic Weapons Research Establishment shock tube is rectangular in cross section with internal dimensions 8 in. × 2 in. and constructed to very close tolerances. It is made up of sections 2 ft. 6 in. long; one of these sections is used as the compression chamber, and the expansion chamber consists of seven sections followed by an observation chamber fitted with windows suitable for the photography of the shock phenomena. Cellulose acetate sheet is used as the diaphragm and this is ruptured by an electric spark.

† In this country Timbrell (1951) has used an electron multiplier cell in conjunction with a Mach–Zehnder interferometer in acoustical studies.

The velocity of the shock wave is measured in the final section of the expansion chamber, before the observation section, by means of light screens which are electronic–optical devices for detecting the passage of a shock wave; they are similar to those described in a paper by Bleakney, Weimer and Fletcher (1949). Briefly the action of a light screen consists of the interception and refraction by a shock wave of a light beam which passes across the tube to a photomultiplier system. Refraction of the light beam increases momentarily the light intensity incident on the photomultiplier, producing an electrical pulse; this pulse is suitably amplified and passed to a crystal-controlled oscilloscope which enables the time intervals between the signals from the light screens to be measured accurately. Any of these signals may, in addition, be passed to a unit which delays it by a variable controlled amount, and it is then used to trigger the light source.

The shock phenomena are normally observed by spark shadow photography, and occasionally by Schlieren photography, using the light from a single air spark of duration about 0·5 microsecond, collimated by a concave mirror. A multiple spark unit, which provides up to six sparks of slightly greater duration and with an adjustable time interval between sparks, has been developed for taking multiple shadowgraphs on one photographic plate. Recently an interferometer of the Mach–Zehnder type has been obtained so that the shock tube is now equipped to give quantitative answers to a large range of shock wave problems.

§ 3. NATURE OF THE EXPERIMENT

The purpose of the present experiment was to investigate the degree by which a plane shock wave of strength about 1·7 is attenuated when it passes through a certain type of wire gauze screen. The experimental method was to clamp the gauze between a special Transpex† observation section and the remainder of the expansion chamber and fire a plane shock through the gauze into the Transpex section where it could be observed by means of spark shadow photography. A diagram of the experimental arrangement is shown in figure 1.

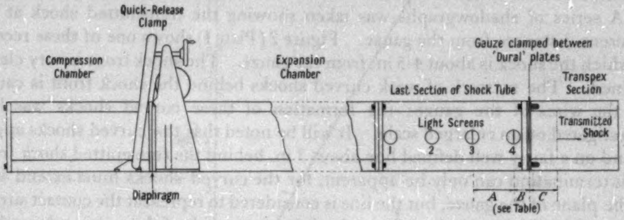


Figure 1.

The strength of the shock incident on the gauze was calculated from its velocity measured by the light screens. Spark shadowgraphs of the transmitted shock were recorded at prescribed times and a distance–time curve for this shock was plotted. The curve was in fact a straight line, the slope of which gave the velocity of the transmitted shock. The strength of the transmitted shock was then calculated from its velocity.

---

† Perspex of high optical quality.

The results were examined in relation to data which were available for the steady flow of air through screens. Finally, the process of transmission of the shock through the gauze was examined in some detail.

§ 4. EXPERIMENTAL DETAILS

The gauze used in this investigation was made of stainless steel wire of diameter $1{\cdot}52 \times 10^{-2}$ in. and the mesh size (distance between centres of adjacent wires) was $5{\cdot}18 \times 10^{-2}$ in. by $5{\cdot}01 \times 10^{-2}$ in. The solidity ratio of the gauze, defined as blocked area divided by total area, was therefore almost exactly 50%. A piece of this gauze 11 in. × $5\frac{1}{4}$ in. was clamped between two similar plates of $\frac{1}{8}$ in. 'Dural' of the same size as the gauze but having a central aperture 8 in. × 2 in., equal to the cross section of the shock tube. The assembly was then bolted between the last section of the shock tube and the Transpex section. This arrangement ensured that the impact of the shock did not seriously deform the gauze; in practice only slight bulging occurred.

The low pressure section of the shock tube was kept at atmospheric pressure ($14{\cdot}47$ lb in^{-2}) and the high pressure section was supplied with compressed air at 35 lb in^{-2} above atmospheric pressure. Diaphragms consisting of two thicknesses of 3-mil cellulose acetate sheet were used. The starting pressures were chosen to produce a shock of strength about 1·7 (overpressure about 10 lb in^{-2}).

When the diaphragm was burst a plane shock travelled down the tube, passed through the gauze and emerged into the Transpex section; there it was observed by means of a conventional spark shadowgraph system. Light from the 0·5 μsec spark was collimated by a 12 in. diameter f/8 parabolic mirror and directed normally through one side of the Transpex section on to a photographic plate held at the other side. The arrangement produced shadowgraphs of natural size. The signals emitted by the third and fourth light screens when the shock passed these stations were recorded on a triple spiral oscilloscope for timing purposes. The signal from the fourth light screen was also used to trip the shadowgraph spark, after a suitable delay had been imposed to record the shock in the desired position, and the instant at which the spark fired was recorded on the oscilloscope.

A series of shadowgraphs was taken showing the transmitted shock at five different distances from the gauze. Figure 2 (Plate I) shows one of these records in which the shock is about 4·5 in. from the gauze. The shock front is very clearly defined. The network of weak curved shocks behind the shock front is caused by the wires of the gauze; the formation of these curved shocks was later investigated on an enlarged scale. It will be noted that the curved shocks appear to end on a fairly well defined line about 3 in. behind the transmitted shock front. This termination can only be apparent, for the curved shocks must extend back to the plane of the gauze, but the line is considered to represent the contact surface between air which was originally on the downstream side of the gauze and air which has passed through the gauze. Turbulence in the transmitted air stream renders the curved shocks invisible on the upstream side of the contact surface. Support for this view is given by Plate V in which the ends of the two innermost curved shocks fail to appear through the highly turbulent air issuing from the shock tube.

The threaded rod which appears on the right of figure 2 (Plate I) was attached to a plate sealing the end of the Transpex section and was used as an index to define the position of the transmitted shock relative to the gauze, which is just outside the field of view to the left of figure 2.

§ 5. Results

Five shadowgraphs of the transmitted shock were taken. The average velocity of the five corresponding incident shocks was found from the light screen measurements to be 1405 ft sec^{-1}. The usual equation relating the strength and velocity of a shock in air is

$$y=\tfrac{1}{6}[7(U/C)^2-1] \qquad \ldots\ldots(1)$$

where y = shock strength (>1), U = shock velocity, C = velocity of sound in air ahead of shock.

In the present experiment the temperature of the air in the shock tube was 19°c and the relative humidity 60%, hence, from the formula for the velocity of sound given in *International Critical Tables*, $C=1125$ ft sec^{-1}. Substituting the values for U and C in equation (1) we find $y=1{\cdot}65$, i.e., the average incident shock strength was 1·65 and the corresponding overpressure 9·4 lb in^{-2}.

The progress of the transmitted shock is recorded in the table and a distance–time curve for this shock is plotted in figure 3. The table was compiled as follows.

Number of shock	1	2	3	4	5
Distance C (in.)	0·74	2·65	4·28	5·93	7·41
Time A (μsec)	456·7	464·1	458·3	464·0	466·4
Time ($B+C$) (μsec)	189·6	308·6	403·5	510·9	605·0
Time B (calculated) (μsec)	145·3	147·6	145·8	147·6	148·3
Time C (calculated) (μsec)	44·3	161·0	257·7	363·3	456·7
Velocity (U) over distance A (ft sec^{-1})	1420	1400	1415	1400	1390
Distance C' (in.) for $U=1405$ ft sec^{-1}	0·73	2·67	4·26	5·97	7·49

For nomenclature see figure 1. Distance $A=7{\cdot}78$ in., distance $B=2{\cdot}47$ in., average velocity U of incident shock = 1405 ft sec^{-1}.

As the distance from the fourth light screen to the gauze (distance B in figure 1) was only about 2·5 in. it was assumed that the measured velocity between the

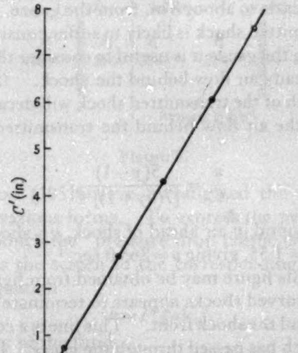


Figure 3. Distance–time graph for transmitted shock. Distance C' = distance C (see figure 1) corrected to incident shock velocity of 1405 ft sec^{-1}.

third and fourth light screens was maintained until the shock struck the gauze. This instant was calculated from the oscilloscope timing records and was taken as zero time in plotting figure 3. The distance of the transmitted shock from the gauze was measured directly from the shadowgraph plate by a travelling microscope. A correction factor (average incident shock velocity) ÷ (actual incident shock velocity) was applied to each measured distance in order to reduce all results to correspond to the average incident shock velocity of 1405 ft sec^{-1}. The scatter of measured velocities about this value was less than $\pm 1\%$ and it was assumed that over this small range the velocities of incident and transmitted shocks were proportional.

Figure 3 consists of the corrected distances plotted against corresponding times. As the points lie on a straight line through the origin it follows that the velocity and strength of the transmitted shock are constant in the region observed and also that the formation of the shock on the downstream side of the gauze takes place with negligible delay.

By means of the method of least squares to fit the best straight line to the points plotted in figure 3, the velocity of the transmitted shock was found to be 1360 ft sec^{-1}. From equation (1), with $C = 1125$ ft sec^{-1} as before, $y = 1{\cdot}55$.

The measured strength of the transmitted shock was therefore 1·55, and the overpressure behind it was 7·9 lb in^{-2}. The accuracy of this last quantity is considered to be to $\pm 0{\cdot}2$ lb in^{-2}.

§ 6. Discussion of Results

The effect of the gauze was to reduce the strength of the shock from 1·65 to 1·55 or, expressed in terms of overpressure, from 9·4 lb in^{-2} to 7·9 lb in^{-2}. Thus, although the gauze blocks about 50% of the cross section of the shock tube it caused only slight attenuation of the shock, at least within the range of observation, which extends to about 8 in. from the gauze. In order to form some idea whether the transmitted shock is likely to suffer considerable attenuation as it progresses further from the gauze it is useful to consider the resistance which such a gauze offers to the steady air flow behind the shock. If this air flow cannot be maintained, the strength of the transmitted shock will decay.

The velocity u of the air flow behind the transmitted shock is given by the equation

$$\frac{u}{C} = \frac{5(y-1)}{[7(6y+1)]^{1/2}} \qquad \ldots\ldots(2)$$

where C = velocity of sound in air ahead of shock, y = shock strength (>1); here $C = 1125$ ft sec^{-1} and $y = 1{\cdot}55$, giving $u = 365$ ft sec^{-1}.

A rough check of this figure may be obtained from figure 2. In this shadowgraph the network of curved shocks appears to terminate on a fairly well defined line about 3·11 in. behind the shock front. This line is a contact surface indicating the front of the air which has passed through the gauze. It originates in the plane of the gauze and then travels downstream at the speed of the air flow behind the transmitted shock. In figure 2 the transmitted shock has travelled 4·28 in. from the gauze in 257·7 μsec (see table). In the same time the contact surface has travelled $4{\cdot}28 - 3{\cdot}11 = 1{\cdot}17$ in. The velocity of the air flow is therefore approximately 381 ft sec^{-1} which is reasonably close to the calculated value of 365 ft sec^{-1}.

The latter will be assumed to be correct. As all the air between the gauze and the transmitted shock is moving with this velocity a corresponding flow must take place through the gauze. This flow will be maintained by the pressure difference which is set up across the gauze by the partial 'head on' reflection of the incident shock, and if the strength of the reflected shock is known this pressure difference can be computed. The strength of the reflected shock was estimated as follows:

Figure 4 (Plate II), which was taken later in the investigation with the gauze mounted at the extreme end of the Transpex section, shows the reflected shock travelling upstream. The apparent distortion of the shock is largely due to optical distortion of the Transpex caused by tapping bolts into it. As the shock appears double, the leading edge was taken to define its position. Two successive shadowgraphs of the reflected shock gave its velocity as 905 ft sec^{-1}, relative to the gauze. When allowance was made for the velocity of the opposing stream of air against which the shock was moving, its strength, calculated from its velocity by equation (1), was found to be 1·23 and the absolute pressure behind it 29·4 lb in^{-2}.

The pressures in the different parts of the system are consequently as shown in figure 5 (*b*). The differential pressure $P_2 - P_3$ across the gauze is $29{\cdot}4 - 22{\cdot}4 = 7{\cdot}0$ lb in^{-2}. It must now be considered whether this is sufficient to maintain the air flow of velocity 365 ft sec^{-1} behind the transmitted shock, and for this purpose we require some data relating to the steady flow of air through a gauze.

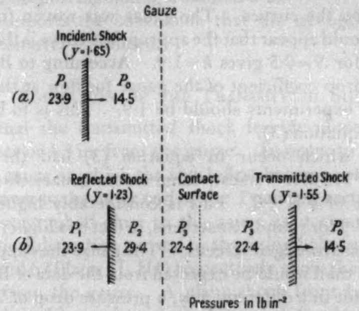


Figure 5.

Baines and Peterson (1951) have investigated the flow of fluids through screens and lattices of various forms. To express the performance of any particular screen they introduce the 'pressure drop coefficient' *k*, which relates the pressure drop *P* across the screen to the corresponding steady stream velocity V_0 according to the equation

$$k = P/\tfrac{1}{2}\rho V_0^2 \qquad \ldots\ldots(3)$$

in which ρ is the density of the fluid.

As Baines and Peterson point out, the dimensionless coefficient *k* depends on the geometry of the screen and on some dimensionless parameter which describes the flow. The appropriate parameter in the present problem is the Reynolds number of the flow. The screen geometry is defined by the solidity ratio *S* and by the form of the screen, i.e. whether it is composed of round or square bars, or woven wire, or is a thin perforated plate. Figure 3 of Baines and Peterson's report shows a set of four curves giving k as a function of S for different screen forms. For constant S the value of k increases, i.e. the resistance of the screen increases, as the elements of which it is composed become increasingly sharp-edged and thinner and consequently produce increasing contraction of the jets through the individual orifices of the screen. Thus, in order of increasing screen resistance the four curves are (*a*) a theoretical curve assuming no jet contraction through the orifices, (*b*) a curve giving experimental data for screens composed of round bars, (*c*) a curve giving experimental data for screens composed of square bars, (*d*) a theoretical curve for a thin plate with sharp-edged holes, supported by some experimental points for perforated plates.

For a solidity ratio of 0·5 the values of k on these four curves are respectively 1·0, 1·9, 3·2 and 3·7. If the value of k for the gauze derived from the pressure drop of 7·0 lb in^{-2} and the corresponding stream velocity of 365 ft sec^{-1} is not less than the appropriate value from Figure 3 of Baines and Peterson's report, then it may be assumed that the pressure drop of 7·0 lb in^{-2} is sufficient to maintain the air flow of 365 ft sec^{-1} behind the transmitted shock, and that the transmitted shock is unlikely to decay through attenuation of the flow behind it.

For all tests to which the four curves refer the Reynolds number R of the flow was in the range 10^3–10^4. Within this range no variation of k with R was found. In the gauze experiment the value of R was about 3×10^3, i.e. it was within the range represented by the curves. The gauze was woven from wire of circular cross section, so it would appear that the appropriate curve is that referring to round bar screens, which for $S = 0{\cdot}5$ gives $k = 1{\cdot}9$. According to Baines and Peterson, then, the pressure drop coefficient of the gauze for flow at the Reynolds number used in the present experiments should be 1·9. This is to be compared with k as calculated from equation (3).

The quantities which occur in equation (3) had the following values: $P = 7{\cdot}0$ lb in^{-2}, $\rho = 0{\cdot}115$ lb ft^{-3} (mean value across gauze), $V = 365$ ft sec^{-1}. The corresponding value of k is 4·2. This is considerably higher than Baines and Peterson's value of 1·9 for round bar screens, in fact it is higher than the theoretical value of 3·7 for thin sharp-edged screens. The gauze thus appears to offer greater resistance to air flow than would be expected from the data of Baines and Peterson. To express this result in a different way, a pressure drop of 7·0 lb in^{-2} across the gauze would be expected to sustain a flow velocity of about 545 ft sec^{-1} whereas the flow is in fact only 365 ft sec^{-1}. It therefore seems reasonable to assume that this velocity is unlikely to diminish with time, and that the observed strength of the transmitted shock is a steady value.

The discrepancy between the two k values of 1·9 and 4·2 is considerable, and it is probable that it is mainly accounted for by over-estimation of the pressure drop across the gauze. This quantity depends partly on measurements made from shadowgraphs which are subject to distortion due to the Transpex walls, so that k value which was calculated as 4·2 may in fact be anywhere between the extreme values of about 2·5 and 5·5. However, it is most unlikely to be as low as the round bar value of 1·9, hence the argument in the previous paragraph is not affected.

Sir William Penney has pointed out that if the passage of the air through the gauze is assumed to be an adiabatic expansion, it is possible to calculate the pressure behind the reflected shock when the incident and transmitted shock strengths are known.

Thus, if the various regions are denoted by the subscripts 0, 1, 2 and 3 as shown in figure 5, and P is the pressure, C the velocity of sound and u the particle velocity in each region, it is known that $P_1/P_0=1{\cdot}65$, $P_3/P_0=1{\cdot}55$, $P_0=14{\cdot}5\ \text{lb in}^{-2}$ and $C_0=1125\ \text{ft sec}^{-1}$.

It follows from the normal shock equations that

$$P_1=23{\cdot}9\ \text{lb in}^{-2} \qquad C_1=1210\ \text{ft sec}^{-1} \qquad u_1=420\ \text{ft sec}^{-1}$$

$$P_3=22{\cdot}4\ \text{lb in}^{-2} \qquad\qquad\qquad\qquad u_3=365\ \text{ft sec}^{-1}.$$

Two equations in P_2 and u_2 may then be set up as follows:

(*a*) Assuming adiabatic expansion through the gauze,

$$P_2 u_2^{\gamma} = P_3 u_3^{\gamma}. \qquad \ldots\ldots(4)$$

(*b*) The particle velocity $u_2=u_1-$(particle velocity due to reflected shock) or

$$u_2=u_1-\frac{5C_1(P_2/P_1-1)}{[7(1+6P_2/P_1)]^{1/2}}. \qquad \ldots\ldots(5)$$

Combining equations (4) and (5) gives

$$P_2\left[u_1-\frac{(5C_1/P_1)P_2-5C_1}{\{7+(42/P_1)P_2\}^{1/2}}\right]^{1\cdot4}=P_3 u_3^{1\cdot4}. \qquad \ldots\ldots(6)$$

The solution of equation (6), obtained by substituting the above values for P_1, P_3, C_1, u_1 and u_3, is $P_2=26{\cdot}8\ \text{lb in}^{-2}$ from which the pressure drop across the gauze is $4{\cdot}4\ \text{lb in}^{-2}$ and the corresponding k value is $3{\cdot}2$.

This calculation supports the conclusion that the strength of the reflected shock has probably been over-estimated.

§ 7. FORMATION OF THE TRANSMITTED SHOCK

Figure 2 shows that the transmitted shock has developed a substantially plane front at a distance of 4·5 in. from the gauze. In order to form some idea of the distance from the gauze at which the shock first becomes plane the gauze was transferred from the inner to the outer end of the Transpex section and the transmitted shock was photographed close to the gauze as it emerged into free air. By this means the optical distortion caused by the fixing bolts in the Transpex was avoided. Figures 4 and 6 (Plates II, III) respectively show the transmitted shock at 1·3 in. and 0·2 in. from the gauze. A plane shock front has formed even at 0·2 in.—a distance of about 13 wire diameters. Figures 2, 4 and 6 all show a network of curved shocks behind the transmitted shock front. The upstream ends of these curved shocks are separated by the same distances as the gauze mesh size so it is clear that they are produced by the individual wires. To obtain a magnified picture of the origin of the curved shocks the gauze was replaced by three horizontal brass rods of diameter about 0·6 in. separated by a distance corresponding to the scaled-up gauze mesh. These rods are seen end-on in figure 7 (Plate IV). This arrangement reproduces a parallel wire grid rather than a woven gauze, but figure 7 clearly shows the shock processes which occur when a plane shock passes through the rods. After travelling approximately two rod diameters the transmitted shock front is still irregular, as sufficient time has not elapsed for the Mach stems in the shadows of the individual rods to overlap.

Finally, a shadowgraph was taken of a plane shock passing through obstacles of intermediate size, formed by six Meccano rods (diameter 0·16 in.) fixed across the end of the Transpex section. Figure 8 (Plate V) shows the development of the curved shocks when the transmitted shock front is approximately 25 rod diameters past the obstacles. At this distance the shock front is almost plane.

§ 8. CONCLUSIONS

A gauze composed of wires of diameter $1{\cdot}52 \times 10^{-2}$ in. and having a mesh of $5{\cdot}18 \times 10^{-2}$ in. by $5{\cdot}01 \times 10^{-2}$ in. offers comparatively little resistance to the passage of a plane shock of strength 1·65 (overpressure 9·4 lb in^{-2}) which strikes it at normal incidence. The transmitted shock has a strength of 1·55 (overpressure 7·9 lb in^{-2}). Within the region observed—up to about 500 wire diameters from the gauze—the strength of the transmitted shock is constant.

Using results obtained by Baines and Peterson for the steady flow of air through screens it appears that the pressure difference across the gauze due to partial 'head on' reflection of the incident shock is sufficient to maintain the air flow required to sustain the transmitted shock.

The region behind the transmitted shock front contains a network of interpenetrating curved shocks which are formed by diffraction of the incident shock by the gauze wires. The transmitted shock front is almost plane after the shock has travelled 13 wire diameters from the gauze.

ACKNOWLEDGMENT

The authors are grateful to the Director, Atomic Weapons Research Establishment, for his valuable comments.

REFERENCES

BAINES, W. D., and PETERSON, E. G., 1951, *Trans. Amer. Soc. Mech. Engrs*, **73,** 467.
BLEAKNEY, W., WEIMER, D. K., and FLETCHER, C. H., 1949, *Rev. Sci. Instrum.*, **20,** 807.
PAYMAN, W., and SHEPHERD, W. C. F., 1941, *Proc. Roy. Soc.* A, **186,** 293.
SMITH, L. G., 1945, *Office of Scientific Research and Development Report* 6271.
TIMBRELL, V., 1951, *Nature, Lond.*, **167,** 306.

The Barkhausen Effect

By R. S. TEBBLE
Department of Physics, The University, Leeds

Communicated by E. C. Stoner; MS. received 6th June 1955 *and in amended form 23rd August* 1955

Abstract. A critical examination is made of the elementary processes which are usually considered as responsible for the discontinuous and irreversible domain boundary movements associated with the Barkhausen effect. The domain structure of an iron-like crystal is considered and it is shown that, of the processes investigated, the only mechanism which will produce discontinuities of the observed order of magnitude and which will result in a hysteresis curve of the usual type, is one in which the movement of a limited section of a 180° domain boundary is delayed by a non-magnetic inclusion whose cross-sectional dimensions are of the same order as the boundary thickness and lie within the range 1 to 2×10^{-5} cm. Inclusions below a critical size ($\sim 10^{-5}$ cm) will be associated only with reversible effects; the energy associated with larger inclusions ($> 10^{-4}$ cm) will be reduced by the presence of subsidiary domain structures and the associated discontinuous boundary movements will be negligibly small. An explanation of the general form of the experimental results is given and it is suggested that any reduction in the hysteresis loss in ferromagnetic materials can best be brought about by the selective reduction of the number of inclusions whose size is within the critical range.

---

§ 1. Introduction

The properties of ferromagnetic materials in relatively low magnetic fields, epitomized in the hysteresis loop, have in recent years been the subject of extensive theoretical and experimental investigation. The means by which the changes in magnetization are brought about are now well understood, not perhaps in detail, but at least in general outline and a clear distinction can be made between domain boundary movements, changes in intrinsic magnetization and the rotation of the domain magnetization vectors. The assessment of the relative contributions of these various processes to the changes in magnetization has been the subject of a considerable amount of work which has been surveyed in a recent review by Stoner (1953).

An important feature of the low field variation of magnetization is that it is made up in part of discontinuous and irreversible changes (the Barkhausen effect), and the magnetization curve is made up of a series of 'steps'. A knowledge of the extent to which the magnetization in particular cases is irreversible is clearly essential to any full understanding of magnetic hysteresis. In the work of Tebble and his collaborators (Bush and Tebble 1948, Tebble *et al.* 1950, 1953), and Newhouse (1952), one of the aims has been to measure the contribution of the Barkhausen discontinuities to the total change in magnetization for a range of materials both in polycrystalline and monocrystalline form. Measurements have been made on annealed and hard drawn iron, large grained iron and single crystals of silicon iron and nickel, involving counting and measuring the size of the Barkhausen discontinuities over the hysteresis cycle. The size of the measured discontinuities covers the range of magnetic moment from about 10^{-7} to 10^{-5} e.m.u. corresponding to volumes of the order 10^{-10} to 10^{-8} cm^3. The number–size distribution curves were found to be monotonic, the number increasing with decrease in size. The only exception to this was the curve for a large grained iron specimen in which a maximum is shown; but it is suspected that this is spurious for in such a specimen the discontinuities occur in bursts and it was difficult to ensure that all the pulses were detected (Tebble, Skidmore and Corner 1950).

Although the experimental distribution curves do not cover the whole size-range of discontinuities, as some of these are too small to be detected, it is exceedingly difficult to imagine any reasonable method of extrapolation to lower values of size which would give a maximum and at the same time satisfy the condition that the integrated contribution from the individual discontinuities should, added to the estimated reversible contribution, give the total change in magnetization. Typical results for the polycrystalline specimens gave a measured Barkhausen contribution of 47% for annealed nickel and 86% for hard drawn iron. The reversible contributions were 5% and 8% respectively. The results for the single crystals were complicated by the effect of shape, for in specimens with large demagnetizing factor N the Barkhausen effect is reduced by a factor of the order of 5 in the specimens examined.

Any satisfactory treatment of the Barkhausen effect must account for the size distributions actually found and explain the apparent absence of any maximum in the distribution curve, account for the effect of demagnetizing fields and relate the size and distribution of the discontinuities with the physical and magnetic properties of the material. Unfortunately the behaviour of specimens which are convenient for experimental examination is seldom amenable to precise quantitative theoretical treatment, and considerable simplification of the problem is necessary. In the present paper the domain configuration of a simple model of a single crystal is investigated and after a consideration of the possible mechanisms by means of which Barkhausen discontinuities could be produced, it is shown that it is only possible to give an explanation of the general character of the experimental results in terms of one specific process.

§ 2. DOMAIN STRUCTURE

A cubic iron-type crystal is considered with directions of easy magnetization in the (100) cube edge direction. The specimen is in the form of a thin rectangular parallelipiped length l, width w, thickness t, whose edges are parallel to [100] directions (figure 1).

The domain structure must be such as to reduce the effect of free poles induced in the specimen, as the energy associated with free poles is relatively large. It is necessary, therefore, that the component of the magnetization normal to the domain boundaries be as far as possible continuous. In figure 1 a simple domain structure for the demagnetized state is shown, and it will be seen that no free poles arise on the 180° boundary. The introduction of a boundary involves an addition to the energy of the system, in that associated with the boundary is a surface energy whose value, depending on the material of the specimen and on the nature

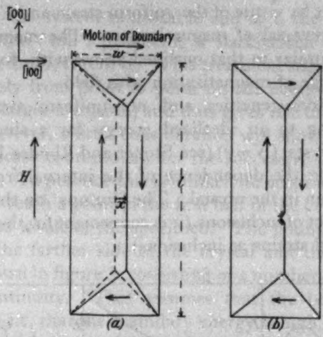


Figure 1. Domain structure of a single crystal of iron, thickness t with each edge parallel to a cubic axis. (*a*) The position of a 180° boundary in a field H. (The position in zero field is shown by the broken line.) (*b*) The retardation of a section of the boundary with the production of free poles.

of the boundary, is of the order 1 erg cm^{-2} (see Lilley 1950). The closure domains at the ends of the specimen form in such a way as to minimize the energy associated with the localized fields due to free poles. A change in magnetization takes place by the movement of the 180° boundary across the crystal to a position, x, as the direction of magnetization along the easy axis is reversed. The alteration in the form of the closure domains gives rise to the formation of free poles with which is associated the demagnetizing energy $\frac{1}{2}NI^2$, where N is the demagnetizing factor and I the average intensity of magnetization $= 2I_0x/w$. The area of the closure domain boundary is, except for specimens with very large demagnetizing factor, small compared with the area of the 180° boundary and the corresponding energy can be neglected. The total energy per unit volume of the system E' in an applied field H is

$$E' = \tfrac{1}{2}NI^2 - HI + E = \frac{2NI_0^2x^2}{w^2} - \frac{2HI_0x}{w} + E \qquad \ldots\ldots(1)$$

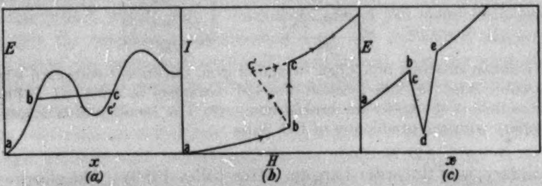
where E is the internal energy per unit volume of the system, in this instance the boundary energy which is sensibly constant once the boundary has been formed. At equilibrium

$$H = NI = 2NI_0x/w.$$

2.1. *The Effect of Strains and Inclusions*

The effect of a uniform strain in the specimen is in general to rotate the magnetization vector away from the easy direction, i.e. the natural anisotropy of the specimen is altered so that the easy direction is no longer parallel to a [100] axis. The direction of magnetization in each domain is then such that if the domain structure remained unaltered, the component of the magnetization normal to the domain boundaries would no longer be continuous and free poles would be produced. In general the boundaries will therefore rotate and be displaced so as to reduce this effect as far as possible. However, the domain structure remains sensibly uniform throughout the specimen, since no region is essentially different from any other by virtue of the uniform strain and E, the internal energy, is constant during a reversal of magnetization. The magnetization process is therefore essentially similar to that outlined above with, for finite values of N, a uniform rate of change of magnetization with field.

The effect of inhomogeneities and non-uniform strain is conveniently considered by referring to an idealized energy for a single boundary in an infinitely long single crystal ($N=0$) (see Stoner and Rhodes 1949), as represented in figure 2 (*a*), showing the dependence of the internal free energy, E, of the boundary, on its position in the crystal. The 'humps' on the curve can be taken as representing the effect of inclusions (it is convenient for the moment to refer to all inhomogeneities and strains as inclusions).



(*a*) (*b*) (*c*)

Figure 2. (*a*) Curve illustrating the variation of internal energy E with boundary position x, associated with continuous and discontinuous boundary movements. From a to b the movement is reversible with $(\partial E/\partial x)_{\max}-2HI_0/w<0$ (see equation (2)). At b, $(\partial E/\partial x)_{\max}=2HI_0/w$, the system becomes unstable and the boundary moves irreversibly and discontinuously to c. (*b*) The corresponding (I, H) curve (see Stoner and Rhodes 1949). In the presence of a demagnetizing field the path taken is bc′ (see Tebble and Newhouse 1953). (*c*) The (E, x) curve associated with a rigid boundary passing an inclusion in an otherwise strain free crystal (see equations (3), (4), (5)).

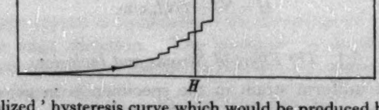


Figure 3. An 'idealized' hysteresis curve which would be produced by a 'stiff' boundary moving across a single crystal.

The position of equilibrium is given by

$$\frac{\partial E'}{\partial x}=\frac{\partial}{\partial x}(E-HI)=0, \quad \frac{\partial E}{\partial x}=\frac{2HI_0}{w} \qquad \ldots\ldots(2)$$

and on the application of a magnetic field the boundary will move to the point on the (E, x) curve where this condition is satisfied. Beyond the point b, where

$2HI_0/w > (\partial E/\partial x)_{\max}$ the system is unstable and at b the boundary moves to a position c, on the next 'hill', producing a Barkhausen discontinuity; this is represented by bc on the (I, H) curve. This treatment suggests that the boundary moves discontinuously from point to point on the energy curve (and from one inclusion to another larger inclusion) and thus gives rise to a series of Barkhausen discontinuities. However, a little consideration will show that this requires that $(\partial E/\partial x)_{\max}$ shall increase continuously for the successive inclusions encountered by a given boundary. In practice this is unlikely to occur and it may well happen that a boundary may reach a point in the first half of the crystal where $(\partial E/\partial x)_{\max}$ is a maximum for the whole of the specimen. The boundary should then move discontinuously to the farther side of the crystal and the magnetization curve would then be as shown in figure 3, consisting of a number of small steps followed by one large discontinuity. This assumes that the boundary moves, or is retarded, as a whole, i.e. that the boundary energy is high, giving rise to stiffness. Although a magnetization curve of this form could no doubt be produced (indeed curves of this general form have been obtained for strained materials) it is clear that it does not occur in the usual materials. Some modification of the concepts behind this approach is needed. In particular, it is necessary to consider the details of the mechanism by means of which the movement of (*a*) a rigid boundary and (*b*) a section of the boundary, is impeded by the presence of a non-magnetic inclusion. It is shown, from a consideration of the effect of the relatively large inclusions in § 3 and of the smaller inclusions in § 4, that only those inclusions whose dimensions are comparable with the boundary thickness can be associated with Barkhausen discontinuities of the observed magnitude.

§ 3. EFFECT OF AN INCLUSION WHOSE WIDTH *a* IS GREATER THAN THE BOUNDARY THICKNESS *b*

3.1. *The Rigid Boundary*

The presence of a non-magnetic material introduces a region of non-uniformity in the magnetization of the specimen and the increase in energy introduced by the presence of free poles may be reduced by the formation of a 'spike' domain structure, as shown in figure 4 (*a*). The effective thickness b of a 180° boundary in iron is given by Lilley (1950) as $b = 1{\cdot}41 \times 10^{-5}$ cm so that a must be at least about 10^{-4} cm, if such a domain structure is to be formed. The length of the spikes and the energy associated with them is to be the subject of another paper and need not be considered in detail here, but it is clear that to have any significant effect on the energy of the system the length of a spike must be of the order $z_0 = 10a$. In the presence of a 180° boundary the energy of the system may be still further reduced when the spike structure is replaced by the subsidiary domain structure shown in figure 4 (*b*). It is assumed that domain structure extends throughout the thickness of the specimens. The boundary energy associated with this structure connecting an inclusion of square cross section, side a, to a boundary at a distance z is given, to a sufficient degree of approximation, by

$$E_1 V = 4\sqrt{2}t\gamma_{90}\,|z| \qquad \ldots\ldots(3)$$

where t is the thickness, V the volume of the specimen and γ_{90} is the energy per unit area associated with a 90° boundary. (Throughout this paper the terms E, E', E_1 etc., represent energy per unit volume.)

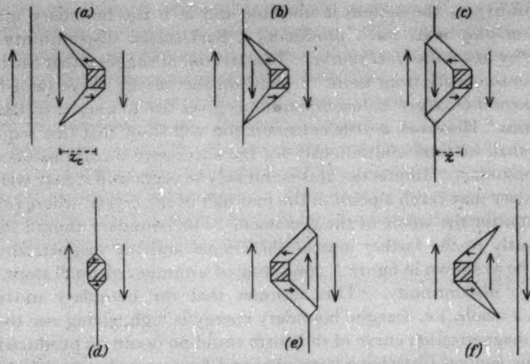


Figure 4. Movement of a 180° boundary past a non-magnetic inclusion. (*a*) Approaching the inclusion, (*b*) just before attachment, (*c*) attached to the inclusion, (*d*) at position of minimum energy, (*e*) just before release, (*f*) after release.

The corresponding energy diagram, shown in figure 2 (*c*), is for a rigid boundary in an iron-like crystal which, except for the presence of inclusions, is free from large strains. The crystal orientation and domain structure are as shown in figure 1. It is assumed that, until the boundary reaches b (figure 2 (*c*)), the internal energy increases smoothly, so that a small finite magnetic field is required to move the boundary. At b the spike domain structure unites with the main boundary (figure 4 (*b*) and (*c*)) and the energy falls to c, beyond which point the energy due to the subsidiary domain pattern is given by equation (3), and the contribution to the total energy of the system, volume V, by

$$E' = E - HI = E_\gamma + 4\sqrt{2t\gamma_{90}}\,|z|/V - HI \qquad \ldots\ldots(4)$$

where E_γ is the energy associated with the 180° boundary.

Thus the energy decreases linearly from c, to a minimum at d. The condition for equilibrium is given by

$$\frac{\partial E'}{\partial x} = 0 = \mp\frac{4\sqrt{2t\gamma_{90}}}{V} - \frac{2HI_0}{w} \qquad \ldots\ldots(5)$$

for z, measured from the centre of the inclusion, negative and positive respectively. Clearly with a boundary to the left of the inclusion (z negative) this condition for equilibrium is never fulfilled for positive fields, and the boundary moves discontinuously to d. It will move on from d when

$$H = H_c = 2\sqrt{2t\gamma_{90}}\,w/I_0V \qquad \ldots\ldots(6)$$

(unless $H > H_c$ when the boundary will move discontinuously across the whole of the inclusion and its associated domain structure). The boundary may break away at any point beyond e, provided that there is some small disturbance capable of displacing the boundary from its position of 'unstable' equilibrium. Whether or not the boundary would then move discontinuously to the next inclusion would depend on the relative magnitudes of the gradient of the (E, x) curve beyond e, and of the applied field.

The boundary movement producing a Barkhausen discontinuity may thus be brought about in one of two ways. In the first the discontinuity is produced as the main boundary unites with the subsidiary domain picture around an inclusion and moves to a position of equilibrium at the inclusion. Clearly the distance moved by the boundary (and hence the size of the discontinuity) will depend on the length of the original spike which is *inter alia* a function of the size of the inclusion. (If H is sufficiently large, the boundary may of course move discontinuously over the inclusion.) In the second process a discontinuity may be produced as the boundary breaks away from an inclusion and, provided $H > (\partial E/\partial x)$ beyond e, moves on to the next inclusion. The size of this discontinuity will thus be a function of the distance between inclusions.

In addition there is the effect of the change in magnetization associated with the disappearance of the spikes as the main boundary unites with the domain structure around an inclusion. The effect of the presence of the spike structure, shown in figure 4 (*a*) and (*f*), is to increase or decrease the magnetization, depending on whether the boundary is to the left or right of the inclusion (z negative or positive). It can easily be shown that the net magnetization of the specimen is sensibly unchanged by the presence of the subsidiary domain structure of figure 4 (*c*), as compared with that of the undistorted plane 180° boundary. The change in domain structure as the main boundary attaches itself to the inclusion, and as it breaks away, is therefore associated with a reduction in magnetization, and with negative discontinuities. These would normally be small (volume $az_c t$) and would be masked by the effect of discontinuities produced by movements of the whole of the 180° boundary (volume $lz_c t$) with which they are coincident: e.g. with $z_c = 10a = 10^{-3}$ cm, $l = 1$ cm, $t = 0{\cdot}1$ cm, $az_c t = 10^{-8}$ cm^3 and $lz_c t = 10^{-4}$ cm^3 compared with the volumes of the measured discontinuities which lie in the range 10^{-10}–10^{-8} cm^3.

Although the shape of the energy curve associated with a single inclusion is considerably different from the corresponding curve in figure 2 (*a*), the type of magnetization curve associated with a general distribution of discontinuities will be similar to that already discussed. The critical field necessary to move a boundary past an inclusion is proportional to a and once the inclusion with the maximum value of a has been passed, the field will be large enough for the boundary to pass all the inclusions which follow and to move uninterruptedly over the remainder of the crystal. A magnetization curve such as that shown in figure 3 could thus be produced.

It seems an inescapable conclusion that the effect of an inclusion on boundary movement is more localized than is assumed in the above treatment. In the following section this treatment is therefore modified to cover the problem of a boundary of which only a limited section is affected by the presence of an inclusion.

3.2 *Effect limited to a Section of the Boundary*

If the displacement of the main 180° boundary is to be limited to a particular segment the production of free poles cannot be avoided. One of the simplest examples is that shown in figure 5 and where a length of boundary is displaced through a distance q, uniformly through the thickness of the specimen. Free poles will be induced as shown, on the displaced section, and the energy of the system will be increased. In addition to the boundary energy due to the subsidiary domain structure around the inclusion (equation (3)), the energy associated with the main 180° boundary will be increased because of the increase in area associated with the displaced segments; there will also be a change in the 'magnetic' energy corresponding to the HI term in equation (2).

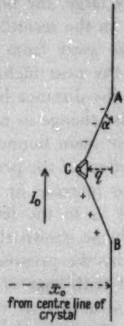


Figure 5. Displacement of a segment AB of the boundary through a distance q to ACB with the production of free poles.

The treatment given below is for an iron-like crystal, as in the previous section: the distance of the undistorted boundary from the centre line of the crystal will be denoted by x_0 and the displacement of the centre of a section, length d, of the boundary by q so that $\alpha = 2q/d$ (figure 5). The following assumptions will be made: (*a*) The effect of an inclusion on the motion by the boundary is mainly limited by the presence of the remaining inclusions to a section whose length d is small compared with the dimensions of the crystal l, t and w; (*b*) the dimensions of the subsidiary domain structure are small compared with d, and $q/d \ll 1$; (*c*) the distorted boundary pattern and the subsidiary domain pattern pass through the thickness of the crystal; (*d*) each Barkhausen discontinuity takes place independently and does not disturb the remainder of the boundary.

The expression for the total energy per unit volume associated with the presence of a displaced section of the 180° boundary is

$$E' = -HI_0 t\, dq/V + E_1 + E_\gamma + E_\sigma \qquad \ldots\ldots(7)$$

where E_1 is the boundary energy due to the subsidiary domain structure set up around the inclusions attached to the boundary and E_γ is the increase in the boundary energy associated with the displaced section of the 180° boundary. As the area of a displaced section will be greater than the corresponding undistorted boundary by a factor $(1 + 2q^2/d^2)$ then, assuming that the boundary energy per unit area remains unaltered,

$$E_\gamma V = 2\gamma_{180} t q^2/d. \qquad \ldots\ldots(8)$$

E_σ is the energy associated with the free poles induced on the displaced segments of the boundary; this term may be also looked upon as representing a change in the demagnetizing energy due to the displacement of a section of the boundary. As the angular displacement α is small the two half sections can be considered as parallel, and the problem is reduced to the evaluation of the energy due to two uniformly charged identical rectangles, lying in the same plane and similarly oriented, and touching. The solution to this type of problem has been given in a comprehensive paper on the demagnetizing energies of magnetized rectangular blocks by Rhodes and Rowlands (1954), and it can be shown that for $d/t \ll 1$ the energy of such a system is, to a sufficient approximation,

$$\sigma^2 d^2 t \ln 2 \qquad \ldots\ldots (9)$$

where σ is the charge density and

$$\sigma = 2I_0\alpha, \qquad \alpha = 2q/d; \qquad \ldots\ldots (10)$$

this gives

$$E_\sigma V = 16LI_0^2tq^2 \qquad \ldots\ldots (11)$$

where V is the volume of the specimen and $L = \ln 2$. The total energy associated with the subsidiary domain structure is thus

$$E'V = 16LI_0^2tq^2 + 4\sqrt{2}\gamma_{90}t\,|z| + 2\gamma_{180}tq^2/d - HI_0qtd \qquad \ldots\ldots (12)$$

where

$$|z| = |x_0 + q - x_z|. \qquad \ldots\ldots (13)$$

The condition for equilibrium is given by

$$\partial E'/\partial q = 0 = 32LI_0^2tq \mp 4\sqrt{2}\gamma_{90}t + 4\gamma_{180}tq/d - HI_0td \qquad \ldots\ldots (14)$$

and the sign of the second term is negative or positive according as z is negative or positive. The above expression gives

$$q_{\text{equil}} = \frac{HI_0d \pm 4\sqrt{2}\gamma_{90}}{32LI_0^2 + 4\gamma_{180}/d}. \qquad \ldots\ldots (15)$$

The second term in the denominator is small and this expression can be written

$$q_{\text{equil}} = \frac{HI_0d \pm 4\sqrt{2}\gamma_{90}}{32LI_0^2}. \qquad \ldots\ldots (16)$$

It will be seen that even without an attached inclusion the boundary will be displaced, over a length d, through a distance

$$\frac{HI_0d}{32LI_0^2} \qquad \ldots\ldots (17)$$

so that on forming the subsidiary domain structure, the boundary segment will move discontinuously over a distance Δq such that

$$\Delta q = 4\sqrt{2}\gamma_{90}/32LI_0^2. \qquad \ldots\ldots (18)$$

Substituting the appropriate values $\gamma_{90} = 1{\cdot}07$ erg cm^{-2}, $I_0 = 1{\cdot}72 \times 10^3$ e.m.u. and $L = \ln 2$ this expression gives $\Delta q = 0{\cdot}92 \times 10^{-7}$ cm. (This may be compared with the thickness of a 180° boundary, $1{\cdot}41 \times 10^{-5}$ cm.) The corresponding volume over which a discontinuous reversal of magnetization takes place is $v = \Delta qtd/2$; taking $t = 0{\cdot}1$ cm, then to give $v = 10^{-10}$ cm^3, d must take the value $1{\cdot}25 \times 10^{-2}$ cm. The production of the larger discontinuities would require d to be of the order of millimetres and would involve the movements of a length of boundary comparable with the width of the crystal. The effect would be similar to the movement of the rigid boundary considered in the previous section.

An additional factor which must be taken into account is the change in magnetization associated with the change from the 'spike' to the pole free subsidiary domain structure of figure 4 (*c*) and vice versa. It has been shown in the previous section that these changes should be associated with a negative discontinuity in the change in magnetization, corresponding to a reversal of magnetization over a volume of the order 10^{-8} cm^3. This would be considerably greater than the effect of the sections of the 180° boundary displacement, which occurs simultaneously, unless the length of boundary involved is again considerable.

Since, in fact, the Barkhausen effect is produced by changes in magnetization in the direction of the magnetic field, it seems unlikely that changes in the subsidiary domain structures about an inclusion can be associated to any appreciable extent with the Barkhausen effect. It must be assumed either that the transition from the 'spike' structure to the pole free domain structure of figure 4 (*c*) takes place continuously or that it rarely occurs.

The reason for the smallness of the boundary movement given by this treatment is the relatively small value of the boundary energy associated with the subsidiary domain structures $(4\sqrt{2}\gamma_{90}t\,|z|)$ compared with energy due to the free pole energy represented by the first term in equation (12). In other words the 'drag' exerted by the subsidiary domain structure is not sufficient to produce any appreciable displacement of a small section of the main boundary; in fact the displacement may be no greater than that produced by the tendency of a section of the boundary to move to the right in the absence of an inclusion (cf. equation (17)).

§ 4. THE EFFECT OF AN INCLUSION, WHOSE WIDTH *a* IS OF THE SAME ORDER OR LESS THAN THAT OF THE BOUNDARY

4.1. *The Rigid Boundary*

It is clear that when the width of the inclusion *a* is approximately equal to that of the boundary *b*, the subsidiary domain pattern discussed above cannot be formed; the boundary will then tend to take up a position such that the energy due to free poles induced on the inclusion will be a minimum as indicated in figure 7. The free pole distribution on the surface of the inclusion will be determined by the direction of magnetization in the boundary (figure 6 (*a*)).

Referring to figure 6 (*a*) and (*b*) it is convenient to express lengths as non-dimensional parameters, e.g. the 'reduced' distance from the centre line of the boundary is ξ, and the actual distance is ξb_0, with b_0 a unit of length, in this instance $1{\cdot}30\times10^{-6}$ cm (Lilley 1950).

It has been shown by Lilley that in a 180° boundary in iron the angle ϕ which the magnetization vector at a distance ξ from the centre of the boundary makes with the easy direction of magnetization is given by

$$\xi = -(1+\tau')^{1/2}\sinh^{-1}[(1-\mu')^{-1/2}\cot\phi] \qquad \ldots\ldots(19)$$

where $\mu' = 1/(1+\tau')$ and τ' is a dimensionless parameter involving the coefficients of elasticity, anisotropy and magnetostriction, and has the value 0·0017 for iron. Curves showing the variation of ϕ with ξ are given by Stoner (1953) and Lilley (1950); using equation (19) Lilley gives a value for the thickness of a 180° boundary in iron as $b = 10{\cdot}87b_0$.

A more manageable expression for ϕ can be obtained from (19) by making the approximation $(1+\tau') = 1$; writing $\alpha = \sqrt{\tau'}$ this gives

$$\cot\phi = \alpha\sinh(-\xi). \qquad \ldots\ldots(20)$$

The 'reduced' free pole density at a point ξ on the surface of an inclusion will then be given by

$$\sigma_\xi = I_0 \cos\phi = \frac{I_0 \alpha \sinh(-\xi)}{(1+\alpha^2 \sinh^2 \xi)^{1/2}}. \qquad \ldots\ldots(21)$$

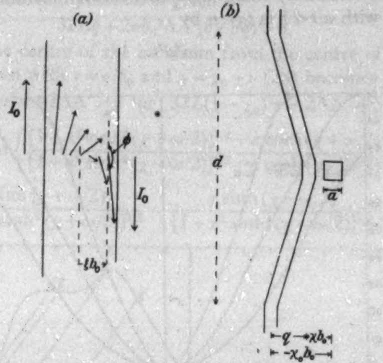


Figure 6. (*a*) Representation of the rotation of the direction of magnetization through the boundary. The origin is the centre line of the boundary; (*b*) displacement of a section of the 180° boundary relative to a small inclusion, $a = \omega b_0$; $q = \nu b_0$.

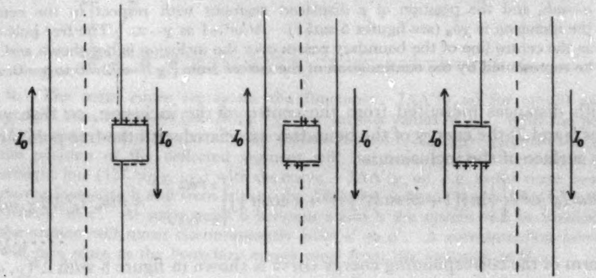


Figure 7. Movement of a 180° boundary over a small inclusion. The position of minimum energy will be at (*b*). The origin for χ is the centre line of the inclusion.

The approximation will be made that the energy associated with the free pole distribution is the same as that due to an equal 'total charge' uniformly distributed over the appropriate surface. The average free pole density will then be

$$\sigma = -\frac{I_0}{\omega} \int_{\xi-\omega/2}^{\xi+\omega/2} \frac{\alpha \sinh \xi \, d\xi}{(1+\alpha^2 \sinh^2 \xi)^{1/2}} \qquad \ldots\ldots(22)$$

where $\omega = a/b_0$, giving

$$\sigma = -(I_0/\omega)\left|\ln\left[(1+\alpha^2\sinh^2\xi)^{1/2}+\alpha\cosh\xi\right]\right|_{\xi-\omega/2}^{\xi+\omega/2} = (I_0/\omega)\Lambda(\xi,\omega). \quad \ldots\ldots(23)$$

It can be shown from the results of Rhodes and Rowlands (1954) that the energy due to two uniformly charged parallel sheets, width a, length t, separated by a distance a, with $a/t \ll 1$ is given by

$$\pi\sigma^2 a^2 t. \quad \ldots\ldots(24)$$

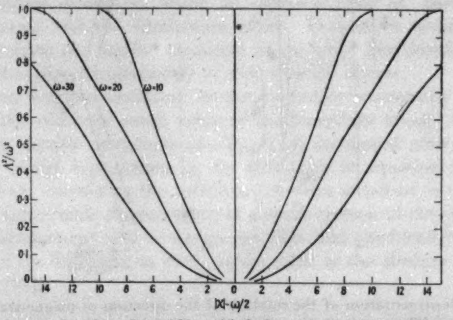


Figure 8. Curves of the function $\Lambda^2(\chi,\omega)/\omega^2$ for $\omega=10$, 20 and 30, plotted against $|\chi|-\omega/2$. The energy associated with the free pole induced on the surface of an inclusion within a boundary is given by $\pi t b_0^2 I_0^2 \Lambda^2$. The width of an inclusion $a=\omega b_0$ and the position of a displaced segment with respect to the centre of the inclusion is χb_0 (see figures 6 and 7). $\Lambda^2/\omega^2 \to 1$ as $\chi \to \infty$. The free pole energy as the centre line of the boundary passes over the inclusion is not shown and would be represented by the continuation of the curves from $|\chi|-\omega/2=0$ to $\chi=0$, $\Lambda^2=0$.

With distances measured from the centre of the inclusion, so that $\chi=-\xi$ (figures 6 and 7) the energy of the boundary associated with the free pole induced on the surface of the inclusion is

$$(\pi t a^2 I_0^2/\omega^2)\left\{\left|\ln\left[(1+\alpha^2\sinh^2\chi)^{1/2}+\alpha\cosh\chi\right]\right|_{\chi-\omega/2}^{\chi+\omega/2}\right\}^2 = \pi t b_0^2 I_0^2 \Lambda^2(\chi,\omega). \quad \ldots\ldots(25)$$

The form of the corresponding energy curve is shown in figure 8 with $\Lambda^2(\chi,\omega)/\omega^2$ plotted against $|\chi|-\omega/2$ the distance of the boundary from the edge of the inclusion. So far as the movement of a rigid boundary is affected, the results will be essentially similar to that discussed in § 3.1; the boundary would move forward to the energy minimum and break away at a critical value of the applied field, which would increase with increasing a and a magnetization curve such as that illustrated in figure 3 could be obtained.

4.2. *Effect limited to a Section of the* 180° *Boundary*

Making similar assumptions to those in § 3.2, with appropriate modifications, the expression for the energy associated with a section of the boundary passing over an inclusion (figure 6 (*b*)) is

$$E'V = 16LI_0^2tq^2 + 2\gamma_{180}tq^2/d - HI_0tqd + \pi t b_0^2 I_0^2 \Lambda^2(\chi,\omega). \quad \ldots\ldots(26)$$

The position of equilibrium is then given by

$$\partial E'/\partial q = 0 = \left(32LI_0{}^2 + \frac{4\gamma_{180}}{d}\right)tq - HI_0td + 2\pi tb_0{}^2I_0{}^2\Lambda\Lambda'\frac{\partial\chi}{\partial q} \quad \ldots\ldots(27)$$

with $\Lambda' = \partial\Lambda/\partial\chi$.

The term $4\gamma_{180}/d$ is small and may be neglected; considering initially small values of H the equilibrium position is given by

$$32Lq + 2\pi b_0{}^2\Lambda\Lambda'(\partial\chi/\partial q) = 0. \quad \ldots\ldots(28)$$

If the distance of the centre of the inclusion from the centre of the undistorted boundary is $\chi_0 b_0$, then with $\nu = q/b_0$ and $\chi = \chi_0 + \nu$ (28) becomes

$$32L\nu/\pi + 2\Lambda\Lambda' = 0 \quad \text{or} \quad 32L(\chi - \chi_0)/\pi + 2\Lambda\Lambda' = 0 \quad \ldots\ldots(29)$$

where

$$\Lambda = \ln\left\{\frac{[1+\alpha^2\sinh^2(\chi+\omega/2)]^{1/2} + \alpha\cosh(\chi+\omega/2)}{[1+\alpha^2\sinh^2(\chi-\omega/2)]^{1/2} + \alpha\cosh(\chi-\omega/2)}\right\} \quad \ldots\ldots(30)$$

and

$$\Lambda' = \frac{\alpha\sinh(\chi+\omega/2)}{[1+\alpha^2\sinh^2(\chi+\omega/2)]^{1/2}} - \frac{\alpha\sinh(\chi-\omega/2)}{[1+\alpha^2\sinh^2(\chi-\omega/2)]^{1/2}}. \quad \ldots\ldots(31)$$

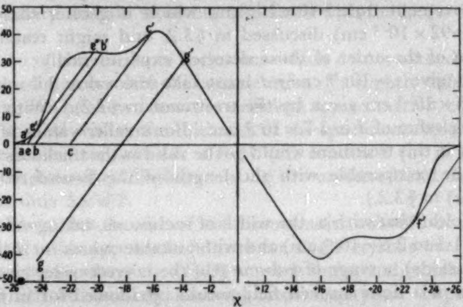


Figure 9. The main curve represents the function $-2\Lambda\Lambda'(\chi, \omega)$ for $\omega = 20$ and the straight lines represent $(32L/\pi)(\chi - \chi_0)$ for various values of χ_0. With the position of the main boundary given by the intersection of $(32L/\pi)(\chi - \chi_0)$ with the χ axis, the position of the deflected segment will be given by the intersection of the straight line $(32L/\pi)(\chi - \chi_0)$ with the curve $-2\Lambda\Lambda'(\chi, \omega)$, e.g. as the main boundary moves from a to b and from b to c, the displaced segment moves from a′ to b′ and from b′ to c′. At some point e between a and b the system will be unstable and the section will move discontinuously from e′ to e″. A corresponding movement will take place as the boundary moves away from the inclusion at f′.

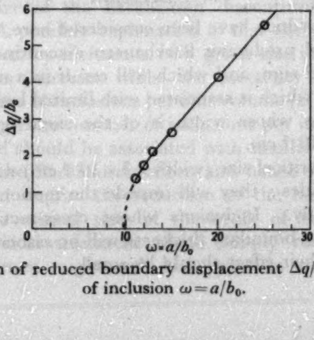


Figure 10. Variation of reduced boundary displacement $\Delta q/b_0$ with reduced width of inclusion $\omega = a/b_0$.

In figure 9 the functions $(-2\Lambda\Lambda')$ and $(32L(\chi-\chi_0)/\pi)$ are shown plotted against χ for $\omega=20$; the position of a section of the boundary when the main boundary is at χ_0 is given by the intersection of the two curves $(-2\Lambda\Lambda')$ and $(32L(\chi-\chi_0)/\pi)$. It is clear that for the particular value of width of inclusion ω chosen $(\omega=20)$, there will be regions of instability near $\chi=-24{\cdot}5$ resulting in a discontinuous boundary movement $\Delta q=4{\cdot}2b_0$, which will be independent of d the length of the boundary segment. The value of $\omega=20$ corresponds to a width of inclusion equal to about twice the width of a 180° boundary and may be taken to represent the probable upper limit to the value which ω can take, because with sensibly thicker inclusions the pole free subsidiary domain structure of figure 4 (*c*) would be produced. When the width of the inclusion is less than a certain critical value, i.e. when ω is such that $\partial(-2\Lambda\Lambda')/\partial\chi<32L/\pi=7{\cdot}06$, there will be no instability and the boundary movement will be continuous. This will occur for $\omega<10$ approximately, as is shown in figure 10 where $\Delta q/b_0$, obtained graphically, is plotted against ω.

The discontinuity near $\chi=-24{\cdot}5$ in figure 9 corresponds to a discontinuous boundary movement $\Delta q\simeq5{\cdot}46\times10^{-6}$cm, which is greater than the boundary movement $(0{\cdot}92\times10^{-7}$ cm) discussed in § 3.2, and might result in changes in magnetization of the order of those detected experimentally; for instance with $t=0{\cdot}1$ cm, to give $v=10^{-10}$ cm^3, d must take the value $3{\cdot}7\times10^{-4}$ cm (cf. the value $d=1{\cdot}25\times10^{-2}$ cm given by the treatment in § 3.2). With $t=10^{-2}$ cm the corresponding value of d is $3{\cdot}7\times10^{-3}$ cm. For smaller values of t, the approximations made in this treatment would not be valid as the thickness of the specimen would then be comparable with the length of the boundary segment α (cf. assumption (*a*) in § 3.2.).

It is thus clear that with a, the width of inclusions, taking values in the range $10b_0$ to $20b_0$ $(1{\cdot}3$ to $2{\cdot}6\times10^{-5}$ cm) and with suitable values for d (i.e. the distance between inclusions), a range of volume v of the correct order could be obtained.

The effect of a finite applied field would be allowed for in figure 9, by the replacement of the straight line bb' by the straight line gg' representing the function $32L(\chi-\chi_0)/\pi-Hd/\pi I_0b_0$, with $bg'=Hd/2LI_0b_0$. The displacement of a section would thus be increased from $\chi_{b'}-\chi_b$ to $\chi_{g'}-\chi_{g}$. The size of a discontinuity would be unaltered however, as the gradient of the straight line is unchanged. The field at which a discontinuity occurs is given by

$$H=[32L(\chi-\chi_0)/\pi+2\Lambda\Lambda']\pi I_0b_0/d. \qquad \ldots\ldots(32)$$

§ 5. DISCUSSION

Of the mechanisms which have been considered here, it seems clear that the only process capable of producing Barkhausen discontinuities of the observed order of magnitude and sign, and which will result in a magnetization curve of the usual shape, is that which is associated with limited lengths of 180° boundary and involves inclusions whose width is of the same order as the boundary thickness, $1{\cdot}3$ to $2{\cdot}6\times10^{-5}$ cm.

Inclusions below a critical size (width $1{\cdot}3\times10^{-5}$ cm) will, on this hypothesis, produce no discontinuities; they will impede the motion of the boundary but will do so continuously. Inclusions whose cross-sectional dimensions are sensibly greater than the boundary thickness will be associated with a subsidiary domain structure and their effect should be small.

The calculated discontinuous boundary movements lie in the range 10^{-6} to 10^{-5} cm and are independent of the length of the boundary affected, within the limits of the assumptions made in the treatment. An approximate relation between the boundary movement Δq and the width of the associated inclusion a is, from figure 10, $\Delta q = Ka$, with $K \sim 3$. The corresponding volumes over which the reversals in magnetization take place, $v = \Delta q t d/2 = Katd/2$, will clearly be dependent on the length of the boundary involved, which is of the order of the spacing between the neighbouring inclusions which delay the boundary. The number of inclusions of a given size a and spacing d will probably increase with decreasing d, so that the smaller discontinuities should be the more numerous —in agreement with the experimental results. An exact estimation of the number and size distribution of the discontinuities associated with volumes which are a function both of inclusion size and spacing, would require considerable information about the structure of a specimen which is normally not available. It is, nevertheless, possible to make some general inferences from the investigation.

The corresponding boundary movements produced by the 'surface tension' effect of the subsidiary domain structure associated with the larger inclusions is of the order 10^{-7} cm.

The application of the results to inclusions which do not extend through the thickness of the specimen involves the consideration of a slightly different domain configuration but it can be shown that, to a sufficient approximation, equations (12) and (26) can be applied with t, the thickness of the specimen, replaced by t' the thickness of the inclusion; $a/t' \ll 1$, $d/t' \ll 1$ (cf. condition (a) in § 3.2). The expressions for the boundary displacement Δq are unaltered and the volume of a discontinuity is thus $\Delta q t' d/2$.

It has been shown that the size of the discontinuities should be independent of the applied field; this appears to be true for single crystals over that part of the magnetization curve where the (I, H) curve is linear and which corresponds to the region in which the boundary is moving across the crystal. In polycrystalline specimens it is to be expected that the results would be in the nature of a mixture of those from each individual crystal grain in the material. The experimental results indicate that the large discontinuities occur near the coercive point, and the reason for this is not at all obvious. If it is assumed that each individual grain is magnetized independently then it would seem that the discontinuities produced in high fields are to be associated with the movement of boundaries across grains with the higher coercivities. If a high coercivity in a particular grain is the result of the greater resistance of the inclusions to the motion of the boundary then the size of the Barkhausen discontinuities should be correspondingly greater. A possible explanation is that the size (volume v) of the discontinuities produced by inclusions of a given size increases as the spacing between the inclusions increases whilst the coercivity might be expected to increase with the number of inclusions per unit volume (cf. equation (32)). Thus those grains with the greater number of inclusions per unit volume would have a higher coercivity and should be associated with smaller discontinuities.

The reduction of the Barkhausen effect, and hence of the hysteresis loss in ferromagnetics, should require the reduction of the linear dimensions of the inclusions to below a critical size, of the order of the boundary thickness ($\sim 10^{-5}$ cm). A physical treatment which simply reduces the size of the impurities in a specimen might well make matters worse, if it results in a reduction of the size of the larger inclusions into the critical range without any compensating reduction in the size of the inclusions at the lower end of the range. This is clearly a point of some importance in the problem of the production of low loss silicon–steel alloys.

§6. CONCLUSION

It is perhaps desirable to emphasize the statement made in the Introduction, on the considerable simplification necessary in problems of this nature (cf. Néel 1946). The model considered in the present paper is that of a thin single crystal of iron with its edges parallel to the [100] cube edge directions and with a single 180° domain boundary passing along the length and through the thickness of the specimen. The effect of closure domains is neglected and it is assumed that inclusions and the associated domain structure extend throughout the thickness of the material.

It has been shown that in such a specimen Barkhausen discontinuities are to be associated with the presence of inclusions whose width is of the same order as the boundary thickness. It is clearly desirable that the treatment should be extended to include the effect of inclusions of various shapes on boundary movements, but a satisfactory appreciation of the results of this work must depend on the extent to which information on the shape and size of inclusions in actual materials is available.

ACKNOWLEDGMENTS

I should like to express my thanks to Professor E. C. Stoner, under whose direction this work has been carried out, and to Dr. V. L. Newhouse, Dr. P. Rhodes, and Mr. G. Rowlands.

REFERENCES

BUSH, H. D., and TEBBLE, R. S., 1948, *Proc. Phys. Soc.*, **60**, 370.
LILLEY, B. A., 1950, *Phil. Mag.* (7), **41**, 792.
NÉEL, L., 1946, *Ann. Univ. Grenoble*, **22**, 299.
NEWHOUSE, V. L., 1952, *Proc. Phys. Soc.* A, **65**, 325.
RHODES, P., and ROWLANDS, G., 1954, *Proc. Leeds Phil. Lit. Soc.*, **6**, 191.
STONER, E. C., 1953, *Rep. Progr. Phys.*, **13**, 83 (London : Physical Society).
STONER, E. C., and RHODES, P., 1949, *Phil. Mag.* (7), **40**, 481.
TEBBLE, R. S., SKIDMORE, I. C., and CORNER, W. D., 1950, *Proc. Phys. Soc.* A, **63**, 739.
TEBBLE, R. S., and NEWHOUSE, V. L., 1953, *Proc. Phys. Soc.* B, **66**, 633.

Magnetic Domain Structure in Twinned Crystals of Silicon Iron

BY R. STREET AND E. O. HALL
Department of Physics, University of Sheffield

MS. received 31*st March* 1955

Abstract. The modification of magnetic domain structures on {110} surfaces of single crystals of silicon iron by the presence of twin bands is examined by colloid deposition techniques. It appears that the structure is determined by the presence of free poles alternately of opposite sign on the interface between the twin and the crystal, giving rise to regular domains separated by 180° boundary walls.

---

§ 1. INTRODUCTION

THE majority of previous work on domain structures, as shown by colloid pattern techniques, has been carried out using large unstrained single crystals. A few investigators, however, have examined the effects of elastic and plastic strains on the domain patterns. Williams, Bozorth and Shockley (1949) have examined colloid patterns on plastically deformed permalloy, and have shown that the colloid concentrates in lines which were considered to be slip lines or lines of incipient cleavage. Dijkstra and Martius (1953) and Dijkstra *et al.* (1954) have examined the effects of stress on the colloid patterns formed on {110} faces of a sample of grain-oriented silicon iron of small grain size, and showed that after plastic deformation the patterns observed were dependent on the state of residual stress.

In the present communication, the influence on the domain structure caused by the presence of twin bands is examined experimentally.

§ 2. SPECIMENS AND METHODS

Investigations were made using grain-oriented $3\frac{1}{2}\%$ silicon iron strips having individual grains of diameter 1 cm approximately. Ideally, the (110) plane is the preferred surface plane, and the [001] direction lies in the rolling direction. The length of the specimen was parallel to the [001] direction, the width was 0·8 in., and the thickness was 0·015 in. The samples had been used previously in an investigation of the fracture of silicon iron crystals by Tipper and Hall (1953) and near the fracture edge, which was the trace of the (001) cleavage plane, a profuse collection of twin bands had appeared. The crystals had broken in a brittle manner so that the grains were otherwise strain free. Figure 1 indicates the form and orientation of the specimens and also shows the (112) and $(11\bar{2})$ twins; these two modes are preferred because they support the maximum shear stress during the tensile test.

Originally, the specimens had been polished on a buffing wheel and even in this state regular colloid deposits were observed on the (110) surface, affected only to a small extent by residual strains. However, all the photographs given here were obtained after the specimens had been polished in a chromic–acetic acid bath (Morris 1949). Although this rounded off the twin boundaries, they were still visible microscopically.

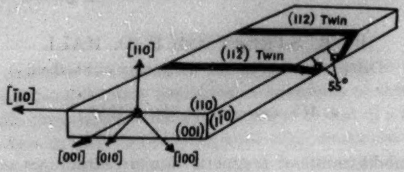


Figure 1. Principal planes and directions in an ideally oriented crystal.

The colloid of Fe_3O_4 in soap solution was prepared as described by Elmore (1938). The colloid patterns were examined by means of a Vickers projection microscope in which the surface to be observed faces vertically downwards. The colloid on the surface was retained and spread uniformly over the crystal by a small piece of microscope cover glass held in position by surface tension forces. Facilities were available for microscopic examination by dark field illumination, which was the method adopted by Bozorth and others, but it was found more satisfactory to observe the patterns here using direct illumination. Magnetic fields of variable magnitude and direction could be applied by means of a small electromagnet, the poles of which were placed in direct contact with the specimen.

§ 3. RESULTS

The general form of the patterns obtained in zero applied field is shown in figure 2 (Plate I). The twin bands are the parallel lines running across the photograph and the edges of the twins are marked in many cases by dotted lines of colloid deposit on both sides of the band. In the regions between the bands the colloid forms into a regular series of lines running parallel to the [001] crystal direction. In general there is little correlation between the lines of colloid formed on the two sides of a twin band. It should be noted here that the crystal faces were not quite (110) planes; a consideration of the crystallography shows that the (112) and $(11\bar{2})$ twins are not quite parallel, nor is the [001] direction exactly perpendicular to the trace of the twin band. However, the domain patterns of the type shown in figure 2 were only found on surfaces close to true {110} faces, i.e. within ± 5°; for example on one grain whose faces were 20° from (110) no such pattern was observed.

Figure 2 is to be compared with figure 3 (Plate I) which shows the colloid deposit very near a fracture edge in a region which is free of twins. In regions which are still free of twins and at distances greater than about 2 mm from the fracture edge, short irregular deposits are obtained which are very similar to those observed on a (110) face by Kaya (1934). The regularity of the deposits between the twin bands and the occurrence of similar though less pronounced patterns near a free edge suggests that they are domains formed to reduce the magnetostatic energy of a free magnetic pole distribution.

The nature of the domain pattern may be elucidated by examination of the patterns obtained when magnetic fields are applied in various directions. Figures 4, 5 and 6 (Plate II) show the appearance of the colloid patterns on the same region of a specimen when a magnetic field is applied in the [001] direction. In practice, since the twins appear very near the fracture edge, it was impossible to avoid a field component perpendicular to the plane of the photograph. Thus in figures 4–6 there is a small vertical component of field intensity. Figure 4 shows the appearance when the applied field intensity is zero; figure 5 was obtained with the field on, and figure 6 with the field reversed. The following points may be noted from these photographs:

(i) With zero applied field the deposition of the colloid is confined mainly to the domain boundary walls. When field is applied there is preferential deposition of colloid on alternate areas between the domain walls. Reversal of the total field, and hence of its vertical component, results in an interchange of the dark and light areas. This behaviour is typical of the behaviour of colloid deposition on a domain system consisting of alternate parallel and antiparallel orientations of magnetic vectors, lying in this case along the [001] and [001] directions.

(ii) On all three photographs, and in particular on those shown in figures 5 and 6, colloid striations are apparent which are perpendicular to the [001] direction. Williams, Bozorth and Shockley (1949) have pointed out that the striations always lie at right angles to the magnetic domain vectors. This observation thus supports the evidence for the structure indicated in (i).

(iii) For zero field, short line deposits on the twin bands lie on a continuous straight line. On the application of the field, these short line deposits are displaced parallel and antiparallel to the [001] field direction and assume a castellated appearance; the sense of the displacements is again dependent on the direction of the applied field. Thus it appears that fields applied along the length of the domains can result in slight lengthwise growth of the domains into the twin bands.

Figures 4–6 illustrate another feature which is commonly observed: the domains have a fine structure consisting of daggers penetrating the main domains whenever the twin bands have an obtuse angle of inclination to the (110) surface of observation (upper twin in these photographs). If the twin band inclines at an acute angle to the surface of observation (lower twin) the fine structure either does not appear or is very much less pronounced.

Figures 7 and 8 (Plate III) show the colloid pattern when a field is applied parallel to the fracture edge, i.e. along the $[1\bar{1}0]$ direction. The region is the same as that used for figures 4–6. This results in a complex breakdown of the pattern which on increasing the field from 20 to 100 Oe progressively spreads to the whole region between the twin bands. Figures 7 and 8 also show a rudimentary structure which is commonly observed both with and without an applied field within the twin band itself. This structure consists of more or less continuous lines parallel to the $[1\bar{1}0]$ direction and has at the moment no adequate explanation.

§ 4. CONCLUSIONS

For the (110) surfaces which have been examined here there is only one easy direction of magnetization (the [001] direction) which lies in the plane of observation. Thus the structure which gives no free poles on this surface consists of domains which are magnetized alternately along the [001] and [00$\bar{1}$] directions. In most cases the domains extend to the twin band and produce free poles at the discontinuity in the crystal structure, resulting in dotted lines of colloid at the twin interface. The dotted lines often appear on both interfaces between the twin band and the crystal, as may be seen for example in the doubled dotted lines of colloid in figure 2. This observation indicates that to a large extent the twin bands, if sufficiently broad, can effectively isolate one part of the crystal from another. Assuming that the observed surface domains magnetized parallel and antiparallel to the [001] direction penetrate throughout the depth of the crystal (cf. Dijkstra *et al.* 1954), then the simple analyses given by Kittel (1949) may be applied. Elementary considerations show that even though in the present case the crystal is divided into oblique prisms the formulae derived by Kittel for right prisms are directly applicable. Two of the cases discussed by Kittel are of interest: (*a*) the case when the crystal is divided by parallel planes into slab domains magnetized alternately in opposite directions; (*b*) the case as in (*a*) but where the magnetostatic energy is made zero by the introduction of triangular prismatic closure domains. These two cases are illustrated in figures 29 (*b*) and 31 respectively of Kittel's paper. For both cases the energy per cm^3 of crystal and the equilibrium distance between the domain boundary walls may be calculated and are

$$\left.\begin{array}{llll} \text{Case }(a) & w_a = 2\left[\dfrac{1{\cdot}7 I^2 \sigma_w}{L}\right]^{1/2} \text{erg cm}^{-3}; & & D_a = \left[\dfrac{\sigma_w L}{1{\cdot}7 I^2}\right]^{1/2} \\ \text{Case }(b) & w_b = \left[\dfrac{2\sigma_w K}{L}\right]^{1/2} \text{erg cm}^{-3}; & & D_b = \left[\dfrac{2\sigma_w L}{K}\right]^{1/2} \end{array}\right\} \quad \ldots\ldots(1)$$

where L is the length of the crystal measured along the [001] direction, K is the anisotropy energy, I is the surface pole density induced on the twin interface and σ_w is the surface energy density of a Bloch wall.

From the observed domain patterns, model (*a*) is applicable, at least as a first approximation, and hence here w_a/w_b should be less than unity. The predicted proportionality of the domain width D_a to $L^{1/2}$ has been tested and the results are plotted in figure 9. The points indicated by circles were obtained from portions of the crystals bounded by relatively broad twin bands. These observations are in moderately good agreement with the predicted relation. Two other points, shown as crosses, are displaced from the straight line and these were derived from the patterns formed on portions of the crystal bounded by relatively narrow twin bands; also in these cases there is a tendency for the boundary walls to be continuous across the twin bands.

The slope of the line in figure 9 gives $I = 120$ c.g.s. units. From equations (1) the ratio $w_a/w_b = (3{\cdot}4)^{1/2} I/K^{1/2}$ and substituting the value of K for silicon iron, given by Kittel (1949) as $2{\cdot}8 \times 10^5$ erg cm^{-3}, $w_a/w_b = 0{\cdot}00132$ which is appreciably less than unity, as expected. The calculated value of I is much less than the spontaneous magnetization of silicon iron at room temperature and it must be assumed that if the model considered is truly applicable, flux closure occurs within the twin bands. The rudimentary structure within the twin bands already noted in § 3 may in fact be due to such closure formations. In cases when a twin band is narrow, it is likely that flux closure is not entirely confined to the twin. Appreciable interaction then occurs between domains on both sides of the twin band, and the effective length L is greater than that measured between successive twin bands. This would explain the anomalous points plotted as crosses in figure 9.

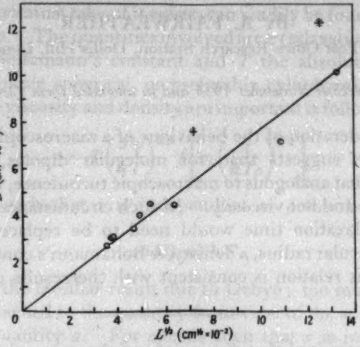


Figure 9. Variation of domain width with domain length.

Acknowledgments

The authors would like to thank Professor Sucksmith for his interest in this work and for the facilities of his department, and Dr. C. F. Tipper for providing the crystals.

References

Dijkstra, L. J., and Martius, U. M., 1953, *Rev. Mod. Phys.*, **25**, 146.
Dijkstra, L. J., Martius, U. M., Chalmers, B., and Cavanagh, P. E., 1954, *Non-destr. Test.*, **12**, (i), 13.
Elmore, W. C., 1938, *Phys. Rev.*, **54**, 1092.
Kaya, S., 1934, *Z. Phys.*, **90**, 551.
Kittel, C., 1949, *Rev. Mod. Phys.*, **21**, 541.
Morris, C. E., 1949, *Metals Progress*, **56**, 696.
Tipper, C. F., and Hall, E. O., 1953, *J. Iron St. Inst.*, **175**, 9.
Williams, H. J., Bozorth, R. M., and Shockley, W., 1949, *Phys. Rev.*, **75**, 155.

Dielectric Absorption in Dilute, Liquid, Polar Solutions: A New Approach

By A. FAIRWEATHER
Post Office Research Station, Dollis Hill, London

MS. received 25*th November* 1954 *and in amended form* 12*th August* 1955

Abstract. Reconsideration of the behaviour of a macroscopic dipole on the lines initiated by Debye suggests that, for molecular dipoles, there may exist a mechanism, somewhat analogous to macroscopic turbulence, for which energy loss is related to density and not viscosity. In such circumstances the familiar Debye formula for the relaxation time would need to be replaced by $\tau \propto (a^5\rho/\boldsymbol{k}T)^{1/2}$ (where a is the molecular radius, ρ density, $\boldsymbol{k}$ Boltzmann's constant, and T absolute temperature). This relation is consistent with the results of experiment in the microwave range.

It is generally recognized that discussions of the properties of polar dielectrics based on supposed analogies between the behaviour of macroscopic and molecular dipoles are not always of fundamental significance. Results obtained by such methods are no more than suggestive and may be misleading. But if supported by experiment they may provide a fresh starting-point for more rigorous analysis: the purpose of the present note is to place on record one such result.

The concept of viscosity, long since abandoned in the theory of solid polar dielectrics, has been of only limited value for liquids. There are circumstances in which a correlation appears to exist between macroscopic viscosity and relaxation time, but there are others in which no such correlation is apparent (Smyth 1950). On the molecular scale this is not surprising. Moreover, even on the macroscopic scale the assumption of a completely viscous constraint cannot always be justified. It is then valid only for dipole movements at 'low' speeds: at much higher speeds, flow is entirely turbulent and energy loss would depend upon density and not viscosity: at intermediate speeds, it would be a function of both viscosity and density.

This suggests that on the molecular scale a somewhat similar situation might perhaps obtain. It might be, as with viscous flow in which a dipole may be supposed to remain attached to its environment, that there is a molecular mechanism analogous to turbulent flow, in which a dipole becomes in some way detached from its environment. This analogy must not be pushed too far: it does not imply that the motion of a molecular dipole is influenced to any significant extent by an external electric field; internal forces inevitably predominate. Polar molecules in solution are in motion with a mean energy of the order $\boldsymbol{k}T$ in every mode, but the way in which they move depends upon the nature of the polar molecule and of its environment. It would not, however, be unreasonable to expect that, in general, small polar molecules in a non-polar medium of low viscosity might move faster, and be associated with shorter relaxation times, than large ones in a medium of high viscosity: and that short relaxation times might correlate with density and long ones with macroscopic viscosity.

An experimental search for a correlation between relaxation time and density would best be undertaken in the microwave frequency range with the aid of dilute solutions of polar solutes in non-polar solvents of low viscosity. The form that such a correlation must take, if it exists, can readily be found by the methods of dimensional analysis. The quantities involved are τ (relaxation time), η (viscosity), ρ (density), $\boldsymbol{k}T$ ($\boldsymbol{k}$ Boltzmann's constant and T the absolute temperature) and a (the radius of a simple spherical, or preferably cylindrical, model of the polar molecule). If both viscosity and density are important it follows that

$$\tau = A\left(\frac{a^5\rho}{\boldsymbol{k}T}\right)^{1/2} f\left[\eta\left(\frac{a}{\boldsymbol{k}T\rho}\right)^{1/2}\right]$$

For dipole speeds such that τ is independent of ρ this becomes, for viscous conditions,

$$\tau = B\eta a^3/\boldsymbol{k}T.$$

With $B = 4\pi$ this is the familiar result due to Debye; too much importance must not, however, be attached to the quantity 4π in view of the approximation implicit in the use of the quantity a. For speeds such that τ is independent of η, the corresponding result for turbulent conditions is

$$\tau = C(a^5\rho/\boldsymbol{k}T)^{1/2}.$$

A, B and C are all dimensionless constants. It is convenient to replace τ by its reciprocal f', the frequency at which $\tan\delta$ (the tangent of the loss angle) is a maximum, $(\tan\delta)_{\max}$. Values of f' relevant to viscous, turbulent or experimental conditions will be identified as $f'_{\text{visc}} \propto T/\eta$, $f'_{\text{turb}} \propto (T/\rho)^{1/2}$ and f'_{exp}.

It follows from the Debye theory that $(\tan\delta)_{\max} \propto 1/T$. This relationship is not affected by the considerations of the preceding paragraphs.

An adequate experimental check of the preceding theory must provide the following evidence. First, a demonstration that conditions exist consistent with $f'_{\text{turb}} \propto (T/\rho)^{1/2}$. Secondly, a confirmation of the view that such conditions have frequently been approached, and not infrequently encountered, in many published measurements; and therefore, that the familiar relation $f'_{\text{visc}} \propto T/\eta$ is not relevant to the interpretation of such measurements. And, thirdly, that the magnitude of $(\tan\delta)_{\max}$ varies as $1/T$.

A quantitative study was undertaken with dilute solutions (1/100 g cm^{-3}) of the same polar solute in each of two non-polar solvents of very different densities. Benzene and carbon tetrachloride, with densities of about 0·9 and 1·6 respectively at 20°C, were employed as solvents: benzophenone was used as the polar solute, partly because some data regarding its behaviour had already been published by Jackson and Powles (1946). The effect of temperature was studied for carbon tetrachloride only and measurement was restricted to two temperatures additional to, and located nearly symmetrically about, room temperature. The range readily available for experiment was limited, of course, by boiling and freezing points. A convenient upper temperature was fixed by the following consideration. Since carbon tetrachloride is more viscous than benzene at room temperature, there is a certain higher temperature (about 44°C, by interpolation) at which the parameter T/η for carbon tetrachloride has the same value as that for benzene at room temperature. At this higher temperature, f'_{visc} for carbon tetrachloride would be identical with that for benzene at room temperature. The corresponding values of f'_{turb} would, however, differ by a calculable amount. The two additional temperatures were fixed, accordingly, at 44°c and −10°c.

An additional, confirmatory, experiment was done for a non-polar solvent of significantly lower density than benzene. *n*-hexane, with a density of about 0·7 at 20°c was chosen for this purpose. This measurement was, to some extent, complementary to that made with carbon tetrachloride at −10°c. It was not unreasonable to anticipate that for hexane the high frequency side of the tan δ–log (frequency) curve might be somewhat higher, relative to the low-frequency side, than for benzene or carbon tetrachloride. Conversely, for carbon tetrachloride at a sufficiently low temperature, the low-frequency side should be the higher. Whether such distortions would be sufficient to introduce a significant discrepancy between f'_{turb} and f'_{exp} was a matter which could only readily be decided by experiment.

The relevant quantities are listed in the table: they are expressed in the usual units, with *f* in kMc/s, and are based on data given in the tables of Kaye and Laby (1948).

Solvent	Parameter	Temperature (°c)				
		−10	20	40	44	50
Benzene	$\eta\times10^{3}$		6·5			
(B)	$(T/\eta)10^{-3}$		45·1			
$\rho=0{\cdot}9$	$(T/\rho)^{1/2}$		18·2			
Carb. Tet.	$\eta\times10^{3}$	16·3	9·7	7·4		6·5
(CT)	$(T/\eta)10^{-3}$	16·1	30·2	42·2	45·1	49·8
$\rho=1{\cdot}6$	$(T/\rho)^{1/2}$	12·9	13·5		15·3	
n-Hexane	$\eta\times10^{3}$		3·2			
(H)	$(T/\eta)10^{-3}$		91·7			
$\rho=0{\cdot}7$	$(T/\rho)^{1/2}$		21·0			
Visc. flow†	f', CT/f', B at 20°c	0·36	0·67		1·0	
	f', CT	2·6*	5·0*		7·4*	
	f', H/f', B at 20°c		2·03			
	f', H		15·0*			
Turb. flow†	f', CT/f', B at 20°c	0·71	0·74		0·84	
	f', CT	5·2*	5·5*		6·2*	
	f', H/f', B at 20°c		1·15			
	f', H		8·5*			

† f', B at 20°c = 7·4.

The calculated values of f', identified by asterisks in the bottom seven lines of the table, are consistent with the experimental results shown in figures 1 and 2. For carbon tetrachloride at −10°c the measured value is somewhat too low: the discrepancy is associated with a marked distortion which can be attributed to the increasing influence of viscosity at the reduced temperature. The shapes of the curves are as expected, and the anticipated linear relationship between $(\tan\delta)_{\max}$ and $1/T$ is confirmed by figure 3.

The measuring technique was, in principle, similar to that developed by Roberts and von Hippel (1946). The form of the apparatus used was, however, very different; in some respects it was novel, and may be made the subject of a future publication. In addition to the precautions usual with measurements of this kind, special care was necessary in order to avoid the possibility of errors peculiar to temperatures in the vicinity of 44 and −10°c. The liquid–air interface was defined by a mica window and significant errors can arise from spurious

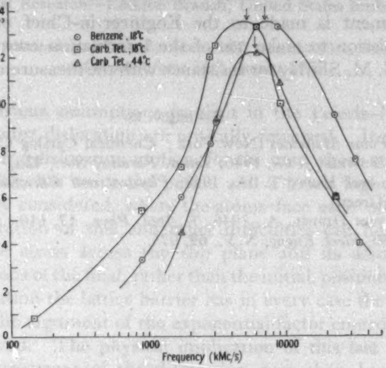


Figure 1.

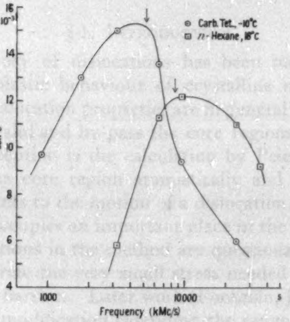


Figure 2.

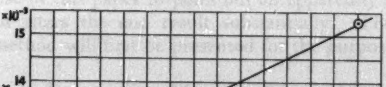


Figure 3.

reflections at this boundary. Such reflections can be produced at the higher temperature, with carbon tetrachloride, by bubbles consisting presumably of a gaseous decomposition product, and at the lower temperature by frost.

ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper, also to Mr. D. G. M. Shirley for assistance with the measurements.

REFERENCES

DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalog Co.), Chap. V.
JACKSON, W., and POWLES, J. G., 1946, *Trans. Faraday Soc.*, **42a,** 101.
KAYE, G. W. C., and LABY, T. H., 1948, *Physical and Chemical Constants* (London : Longmans Green).
ROBERTS, S., and VON HIPPEL, A., 1946, *J. Appl. Phys.*, **17,** 610.
SMYTH, C. P., 1950, *Elect. Engng, N.Y.*, **69,** 975.

Modification of the Peierls–Nabarro Model for Edge Dislocation Core

By H. B. HUNTINGTON †

Office of Naval Research—London Branch, United States Embassy, London

Communicated by M. A. Jaswon; MS. received 13*th June* 1955
and in final form 24*th August* 1955

Abstract. The various assumptions implicit in the Peierls–Nabarro treatment of the core of an edge dislocation are critically reviewed. It is pointed out that the lattice barrier for dislocation motion, for the case where atom rows alternate along the slip plane, has twice the period and a much higher amplitude than for the case usually considered, where the atoms face each other across the slip plane. The resolution of this and other difficulties can be accomplished by arranging that the stress across the slip plane and its associated interaction potential be functions of the final, rather than the initial, position of each atom row. With this modification the lattice barrier has in every case the periodicity of the slip distance but the argument of the exponential factor controlling the height of the barrier is halved. The physical implication of this last result must await a more detailed treatment of the dislocation core than has been previously attempted.

§ 1. Introduction

Though the theory of dislocations has been increasingly successful in explaining the plastic behaviour of crystalline matter, the quantitative calculations on dislocation properties are in general limited to those aspects where elastic theory is valid and by-pass the core regions where large strains are encountered. One exception is the calculation by Peierls (1940) and Nabarro (1947) which treats the core region atomistically and estimates the potential barrier the lattice presents to the motion of a dislocation. Because of its unique position this estimate occupies an important place in the literature and, although some of the approximations in the method are questionable, the result has been widely quoted to illustrate the very small stress needed to keep the dislocation moving over the lattice barrier. Later work (Foreman, Jaswon and Wood 1951) has shown by suitable modification extending the range of the edge dislocation in the slip plane that the lattice barrier can be decreased to an even greater degree.

It is the purpose of this paper to point out an apparently slight variation in the method which alters the end result substantially. The elements of the Peierls–Nabarro method will first be presented for the purpose of reference.

§ 2. Barrier to Dislocation Motion—Peierls–Nabarro Model

Assume a simple rectangular lattice with interatomic distance b along the slip plane, taken to be in the x direction, and with interatomic distance a along the y axis at right angles to the slip plane, so that a is the interslip-plane distance. As a first step to forming the dislocation break the lattice at a slip plane and rejoin after a displacement of $\frac{1}{2}b$. Let now the upper half undergo a compressive displacement $u_+(x)$ and the lower half an expansive displacement $u_-(x)$ so that registry is re-established at large distances, or $u_+(\infty) = -u_-(\infty) = b/4$. One defines a relative displacement function,

---

† Now at Physics Department, Rensselaer Polytechnic Institute, Troy, New York.

$$\phi = \tfrac{1}{2}b + u_+ - u_- \simeq \tfrac{1}{2}b + 2u_+(x - \alpha b) \qquad \ldots\ldots(1)$$

where αb measures the position of the centre of the dislocation from a half plane of atoms above the slip plane perpendicular to the slip direction. So far two possibly critical assumptions have been made: (1) displacements in the y-direction can be neglected and (2) u_+ equals $-u_-$ even in the region of large strains. Two more are involved in the next equation which establishes the shear stress p_{xy} felt by any row of atoms bordering the slip plane as a result of the material on the other side,

$$p_{xy} = -\frac{b\mu}{2\pi a} \sin \frac{2\pi\phi}{b} \qquad \ldots\ldots(2)$$

where μ is the shear constant of the lattice. The additional assumptions are: (3) the shear stress, defined as a function of ϕ from equation (1), depends only on the magnitude of u and not on its direction and (4) the dependence is sinusoidal. It has been shown by Eshelby (1949) and Leibfried and Lücke (1949) on the basis of elastic theory that p and u are related,

$$p_{xy} = \frac{\mu}{\pi(1-\sigma)} \int_{-\infty}^{\infty} \frac{du(s)}{ds} \frac{ds}{s-x} \qquad \ldots\ldots(3)$$

where σ is Poisson's ratio and u is written for u_+. To use this relation as valid in the core region constitutes an additional fifth assumption. An appropriate solution of the above equations turns out to be

$$u(x) = -\frac{b}{2\pi} \tan^{-1} \left\{ \frac{2(1-\sigma)(x-\alpha b)}{a} \right\} \qquad \ldots\ldots(4)$$

for u.

Investigation has shown that the part of the dislocation energy which depends on dislocation position, αb, arises from the interaction of atoms on opposite sides of the slip plane. By integrating equation (2) one establishes an interaction potential per pair of atom rows per unit length of dislocation,

$$V(x) = \frac{b^3\mu}{4\pi^2 a} \left(1 + \cos \frac{2\pi\phi}{b} \right). \qquad \ldots\ldots(5)$$

For the dislocation energy one takes one half the sum of the values of this potential evaluated at the positions of each of the atom rows adjoining the slip plane. Above the slip plane these rows have $x = nb$ and below the slip plane $x = (n + \frac{1}{2})b$. Accordingly for the dislocation energy ($n = \frac{1}{2}m$) one obtains

$$\begin{aligned} E &= \left(\frac{b^3\mu}{8\pi^2 a} \right) \sum_{m=-\infty}^{\infty} \left[1 + \cos 2 \left\{ \tan^{-1} \frac{2(1-\sigma)(\frac{1}{2}m - \alpha)b}{a} \right\} \right] \\ &= \left(\frac{b^3\mu\gamma^2}{\pi^2 a} \right) \sum_{m=-\infty}^{\infty} [4\gamma^2 + (m - 2\alpha)^2]^{-1} \qquad \ldots\ldots(6) \end{aligned}$$

where γ is $a/2b(1-\sigma)$. The summation can be evaluated by applying Cauchy's theorem to the contour integral below

$$\oint_{R\to\infty} \frac{\cot \pi z \, dz}{(2\alpha - z)^2 + 4\gamma^2} = 0. \qquad \ldots\ldots(7)$$

The final result is

$$E \simeq \frac{b^2\mu}{4\pi(1-\sigma)}[1+2\cos 4\pi\alpha \exp(-4\pi\gamma)]. \qquad \ldots\ldots(8)$$

As a result the minimum shear stress to keep a dislocation moving is

$$T_{\mathrm{m}} \simeq \frac{2\mu}{(1-\sigma)}\exp(-4\pi\gamma). \qquad \ldots\ldots(9)$$

The result for the barrier potential, equation (8), shows one apparent peculiarity, namely that the period of the potential is half the Burgers vector. Since it is almost certain that dislocation barriers must in general exhibit the periodicity of the Burgers vector, one looks for the feature of the model which gives rise to this behaviour in this case. The reason the dislocation has the same energy when its centre lies along a half plane in the compressed medium ($\alpha=0$) as it does when it lies on a half plane in the extended medium ($\alpha=\frac{1}{2}$) is that one configuration can be obtained from the other by reflection in the slip-plane and reversal of the signs of all displacements u. This reversal does not alter the dislocation energy since, as has already been pointed out in assumption (3), the model does not distinguish between the directions of u in determining p_{xy} and hence E. Perhaps the easiest way to disturb the peculiar symmetry of the situation under study is to go to the case where the atoms that border the slip plane no longer face each other in the undisturbed lattice but alternate in their position. After the first step in forming the dislocation, namely the relative displacement of $\frac{1}{2}b$, the atom rows would again face each other at regular intervals of b along the slip plane. (Of course a dislocation of this sort corresponds to the removal of two half planes of atoms.) The summation appearing in equation (6) is now a single summation over the integral values of n instead of half the summation over integral and half integral values. The resulting expression for E is

$$E = \frac{b^2\mu}{4\pi(1-\sigma)}[1+2\cos 2\pi\alpha \exp(-2\pi\gamma)]. \qquad \ldots\ldots(10)$$

Here the barrier has the periodicity of the lattice but also the quantity in the exponent has been halved. That the ratio T_{m}/μ should be so greatly altered by a change in atom arrangement causes one to examine the model more critically.

According to the Peierls–Nabarro model the trans-planar stress p_{xy} is a function of the quantity ϕ which depends only on u and is in turn a function only of x the position of the ion row before the u-displacement. Presumably this construction was a natural outgrowth of the original atomic arrangement where the trans-planar stress was visualized in terms of interactions between pairs of atom rows originally facing each other. It has the awkward consequence that the trans-planar stress as a function of position before the u-displacement is not a continuous function of final position. For example there would be a discontinuity in p_{xy} across the slip plane for atom rows that might come to opposing positions from different initial x. The same sort of difficulty is encountered with the interaction potential $V(x)$ which would be the same for two initially opposing atom rows even though one row moved in toward the dislocation centre while the other was displaced outward. For an elastic model where the displacements are small these objections are unimportant but the solution of the Peierls–Nabarro dislocation is narrow and shows the importance of large strains. These difficulties can be avoided if we take ϕ to be the same function of the final position $x' = x + u$ as it is of x according to equation (1). With this revision both p_{xy} and V become continuous functions of the atom row's final position and represent not a pairwise interaction, but one between the row and the whole opposing slip-plane surface.

§ 3. MODIFIED PEIERLS–NABARRO MODEL

According to the proposed modification

$$p_{xy} = -\left(\frac{b\mu}{2\pi a}\right) \sin\left\{\pi + 2\tan^{-1}\frac{2(1-\sigma)(x'-\alpha b)}{a}\right\} \quad \ldots\ldots(11)$$

where $x' = x + u$. Since expression (4) for u was derived from elastic theory, we take its form here to be unaltered. With the simple rectangular lattice considered earlier the evaluation of E involves two separate summations, for the atom rows above and below the slip plane respectively:

$$E = \left(\frac{b^3\mu\gamma^2}{4\pi^2 a}\right)\left\{\sum_{n=-\infty}^{\infty}\left[\gamma^2 + \left(\frac{x_+'}{b-\alpha}\right)^2\right]^{-1} + \sum_{n=-\infty}^{\infty}\left[\gamma^2 + \left(\frac{x_-'}{b-\alpha}\right)^2\right]^{-1}\right\} \quad \ldots\ldots(12)$$

where $x_+' = nb + u(n-\alpha)$, ...(13a) $x_-' = (n+\tfrac{1}{2})b - u(n+\tfrac{1}{2}-\alpha)$ ......(13b)

and $$u(z) = -\left(\frac{b}{2\pi}\right)\tan^{-1}\left(\frac{z}{\gamma}\right). \quad \ldots\ldots(13c)$$

The summations cannot be evaluated as before because the function representing the displacement is not single-valued in the complex plane. The usual methods using contour integrals cannot be used effectively.† A procedure which we have followed to circumvent this difficulty and to gain an approximate evaluation is to replace the correct displacement function $-(b/2\pi)\tan^{-1}(z/\gamma)$ by another having the same value and slope at zero and infinity,

$$u(z) = -\frac{b}{4}\tanh\frac{2z}{\pi\gamma} \quad \ldots\ldots(14)$$

where the ratio of the new function to the original is always less than 1·23. The preceding methods can now be re-applied, in theory at least, to evaluate equation (12). It can be shown that both of the sums give rise to expressions of period b and that the same is true of E.

In practice the zeros for the denominators include complex as well as pure imaginary numbers which complicates the evaluation of the residues, particularly for arbitrary values of γ. To avoid this difficulty we have chosen to work only with the amplitude of the potential barrier that the lattice presents to the dislocations,

$$B = E(\alpha=0) - E(\alpha=\tfrac{1}{2}). \quad \ldots\ldots(15)$$

A second approximation introduced here to render the calculation manageable is to expand both square brackets of equation (12) in powers of u/x and to discard the quadratic and higher terms. This device is of somewhat questionable accuracy but serves in a qualitative way to demonstrate the effect of the revision in the model.

$$E(0) - E(\tfrac{1}{2}) \simeq \frac{b^3\mu\gamma^2}{8\pi^2 a}\sum_{m=-\infty}^{\infty}(-1)^m\frac{m\tanh(m/\pi\gamma)}{[(m/2)^2+\gamma^2]^2} \quad \ldots\ldots(16)$$

where $m = 2n$.

---

† The author is indebted to Mr. A. J. Foreman for pointing out this difficulty.

This sum can be evaluated from the contour integral

$$\frac{1}{2\pi i}\oint_{R\to\infty}\frac{z\tanh(z/\pi\gamma)\operatorname{cosec}\pi z}{[(z/2)^2+\gamma^2]^2}\,dz=0. \qquad \ldots\ldots(17)$$

Besides the residues at $z=m$ there are also residues at $z=i(s+\tfrac{1}{2})\pi^2\gamma$, which are all negligibly small, and residues at $z=\pm 2i\gamma$. One obtains

$$\pi^{-1}\sum(-1)^m\frac{m\tanh(m/\pi\gamma)}{[(m/2)^2+\gamma^2]^2}=\gamma^{-1}\operatorname{cosech}(2\gamma\pi)\,[\pi\tan(2/\pi)\coth(2\pi\gamma)$$

$$+(\gamma\pi)^{-1}\sec^2(2/\pi)]\simeq\gamma^{-1}\exp(-2\gamma\pi)\,[2{\cdot}3+(0{\cdot}49/\gamma)]. \qquad \ldots\ldots(18)$$

It follows

$$B\simeq\frac{b^2\mu}{16(1-\sigma)}\exp(-2\gamma\pi)\,[2{\cdot}3+(0{\cdot}49/\gamma)]. \qquad \ldots\ldots(19)$$

On the basis of an assumed sinusoidal potential $E(\alpha)$ one obtains

$$T_{\mathrm{m}}/\mu=[16(1-\sigma)]^{-1}\exp(-2\gamma\pi)\,[2{\cdot}3+(0{\cdot}49/\gamma)]. \qquad \ldots\ldots(20)$$

In the table are listed the approximate formulae for T_{m}/μ for the various cases considered here and their numerical values for $\gamma=0{\cdot}5$ and $0{\cdot}7$, corresponding to $\sigma=0{\cdot}3$ and a/b equal to 0·7 and 1·0 respectively. The symbol in the first column denotes the model used, whether Peierls–Nabarro (P–N) or modified as here (M). In the next column is denoted the atomic arrangement, whether the atoms face each other (F) or alternate along the slip plane (A). There are no entries for the case M and A because to the approximation in which one considers only the term linear in u/x in the expansion of equation (15) the effect of the modification drops out and the results are identical with those for P–N and A.

Model	Arrangement of atoms	Formula for T_{m}/μ	T_{m}/μ	
			$\gamma=0{\cdot}5$	$\gamma=0{\cdot}7$
P–N	F	$2(1-\sigma)^{-1}\exp(-4\pi\gamma)$	$0{\cdot}005_3$	$0{\cdot}0004_4$
P–N	A	$(1-\sigma)^{-1}\exp(-2\pi\gamma)$	$0{\cdot}06_1$	0·017
M	F	$(1-\sigma)^{-1}\exp(-2\pi\gamma)$ $\times[0{\cdot}15+(0{\cdot}030/\gamma)]$	0·013	0·0032

One feels intuitively that the choice of atomic arrangement should not have a profound effect on T_{m}/μ. Since the use of the modified model is to reduce greatly the disparity between T_{m}/μ for these two cases (for $\gamma=0{\cdot}7$ the ratio goes from 40 to 5·4), this is an additional argument for its validity (and also for the validity of the two approximations used in developing equation (19) from equation (12)).

The most interesting implication of this result with the modified model is that the negative quantity appearing in the exponent of the original expression for T_{m} is too large by a factor of 2. It appears that the same arguments that apply here apply also for the modification considered by Foreman, Jaswon and Wood, since the same summations are involved for arrangements F and A. Adapting their formula (21) one obtains in a corresponding way

$$T_{\mathrm{m}}=\frac{\mu}{1-\sigma}\exp(-2\pi\gamma a_0)\left(1+\frac{K}{2}+\frac{K^2}{24}\right) \qquad \ldots\ldots(21)$$

where $K=2\pi(a_0-1)(a/b)(1-\sigma)$ and a_0 is the parameter a used by Foreman, Jaswon and Wood to extend the dislocation. ($a_0=1$ reduces to the original

P–N dislocation.) If one assumes that T_m/μ is of the order of 10^{-6} or smaller to fit with observed values of the critical shear stress, then on the basis of equation (21) above the lower limit on a_0 will come at 4 instead of 2 as estimated from Table 2 of Foreman, Jaswon and Wood. Also the minimum width of such a mobile dislocation must be greater than $3b$.

§ 4. DISCUSSION

Though it has always been recognized that calculations of dislocation mobility were very sensitive to model, it has not perhaps been realized to what extent the predictions of the most frequently employed model depended on arrangement of atoms along the dislocation plane. The revised model proposed here appears to be preferable both from the logic of its construction and the consistency of its results, but the value of μ/T_m is far smaller than given by the original theory.

It does not follow necessarily that real dislocations actually require such large stresses to surmount the lattice barrier since the analysis of Foreman, Jaswon and Wood (1951) has shown that this barrier drops very rapidly as the dislocation is extended. Very little has been done as yet to determine from theoretical models the atomic distribution of dislocation cores. The evaluation of the lattice potential for the dislocation hinges on a knowledge of the dislocation width and possibly other parameters measuring atomic relaxation. Even if it should be established that the stress to move dislocation lying along crystallographic directions is greater than the minimum critical shear stress, there is still the possibility that dislocation segments lying at an angle to the crystallographic directions may move first, as Seitz has suggested, liberating enough energy to pull the other segments over the barrier.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. M. A. Jaswon, Mr. D. B. Dove and Mr. A. J. Foreman for their very helpful interest in this work.

REFERENCES

ESHELBY, J. D., 1949, *Phil. Mag.*, **40,** 903.

FOREMAN, A. J., JASWON, M. A., and WOOD, J. K., 1951, *Proc. Phys. Soc.* A, **64,** 156.

LEIBFRIED, G., and LÜCKE, K., 1949, *Z. Phys.*, **126,** 450.

NABARRO, F. R. N., 1947, *Proc. Phys. Soc.*, **59,** 256; 1952, *Advances in Physics* (*Phil. Mag.* Suppl.), **1**, 269.

PEIERLS, R. E., 1940, *Proc. Phys. Soc.*, **52,** 34.

Inter-atomic Force Law and Equation of State of Solids

By B. DAYAL† and R. S. SHARMA‡

† Banaras Hindu University, India, and Birkbeck College, London
‡ Banaras Hindu University, India, and Royal Holloway College, London

Communicated by R. Fürth; MS. received 30th March 1955
and in amended form 18th July 1955

Abstract. The pressure of a solid has been divided into two terms p_1 and p_2, of which the first is the static part due to a non-vibrating lattice at $T=0$, and the second arises from the thermal vibrations. For all volumes measured at zero pressure the two are equal and opposite to each other. The static pressures of eight solid elements, aluminium, copper, silver, gold, lead, platinum, tungsten and molybdenum, have been obtained from their thermal pressures $\bar{\gamma}E/V$ by the use of the above condition. The average Grüneisen constant $\bar{\gamma}$ has been calculated in the usual way from the thermal expansion and the specific heat data. These static pressures have been compared with the derivatives of the potential energy as derived from the inter-atomic potential used earlier by Fürth, and a very close agreement has been observed between them.

In his extensive study of the equation of state of solids Fürth (1945) has made use of the equation of state which was derived from rigorous lattice theory by Born and Bradburn (1943) and Bradburn (1943) for elements with a face-centred cubic lattice under the assumption that the inter-atomic potential energy function has the form of a simple power law. He pointed out that, in spite of the fact that the actual inter-atomic force law was certainly of a completely different nature, a fictitious force law of the form mentioned can be used, provided that it resembles the real one in some general features and can be made to fit it in a narrow region surrounding the equilibrium by a suitable choice of the constants. He derived the exponents m and n of his force law from experimental values of the thermal expansion and the compressibility and of their variations with temperature and pressure, and in this way he was able to interpret most of the thermodynamic properties of a large number of solid elements on the basis of the theory.

There are, however, two important limitations imposed in this work. In the first place classical statistics has been used, so that the equation of state is valid only above the Debye characteristic temperature. Secondly the values of the thermal expansion coefficient which have been used in the derivation of the exponents do not represent actual experimental observations but are the extrapolated values which a solid would possess at the absolute zero of temperature if it would obey classical statistics throughout. In view of the importance of the subject it would be an advantage if the law of force could be based on those properties of solids for which a large amount of accurate data is available. It would also be desirable that this be tested over a wide range of temperatures, not necessarily above the Debye temperature.

It was suggested by one of us (Dayal 1950) that the specific heat and thermal expansion data could be used directly to derive and check the form of the potential energy function of crystalline solids. Since these properties have been the subject of extensive experimental studies and data of a very high order of accuracy are available we have used them to test Fürth's power law and to derive more accurate values of the exponents m and n, wherever necessary.

We take the following equation of state due to Grüneisen (1926) as our starting point.

$$\left.\begin{aligned} p &= -\left(\frac{\partial \Phi}{dV}\right)_T + \bar{\gamma} E/V \\ &= p_1 + p_2. \end{aligned}\right\} \qquad \ldots\ldots(1)$$

Here $\bar{\gamma}$ is the well-known Grüneisen constant, $-\partial \log \nu/\partial \log V$. We have put the bar at the top to emphasize that this is not necessarily constant for all the vibration frequencies of a solid as postulated by Grüneisen, but that it is a sort of an average quantity of the same type as the well known Debye characteristic temperature. E is the thermal energy and Φ is the potential energy of the lattice. The effect of temperature is included in the second term p_2 but the first term p_1 is independent of it and is equal to the static pressure at $T=0$. The equation (1) has been directly used in the derivation of the following well-known Grüneisen expression which gives the thermal expansion coefficient α of a solid in terms of its specific heat C_v and its compressibility κ_0 at $T=0$:

$$\alpha = \frac{C_v \bar{\gamma} \kappa_0}{V_0(1 - kE\bar{\gamma}\kappa_0/V_0)^2}. \qquad \ldots\ldots(2)$$

This equation has been very widely used by the experimental workers to interpret their measurements of the thermal expansion coefficients. Here V_0 is the volume of the solid at $T=0$, and k is a constant whose value was found by Grüneisen to lie between 2 and 3. He showed that it is equal to $\frac{1}{6}(m+n+3)$ where m and n are the exponents of the power law used by him. Its exact value is not very important at low temperatures on account of its occurrence in a small correction term. Experimental workers have generally used an empirical value for it, which fits best with their experimental data.

The procedure adopted by us is to go in the reverse order from equation (2) to (1). $\bar{\gamma}$ is calculated from equation (2) using the experimental values of C_v, α and κ_0. This is then substituted in (1) to get the second term p_2. The volumes of the solids at different temperatures are then calculated with the help of the thermal expansion data. Since the experimental measurements are always made at atmospheric pressure which is practically zero in the scale of pressures with which we are concerned, these volumes are effectively volumes at $p=0$. For these volumes, therefore, we can put $p=0$ in equation (1) and we get $p_1 = -p_2 = -\bar{\gamma}E/V$. These values of p_1 are then compared directly with the pressures obtained from Fürth's equation of state for $T=0$. As no term involving temperature is present no particular statistics are involved and Fürth's law of force can be verified and its exponents checked without any limitation.

In these calculations we have used experimental values of the thermal expansion coefficients found by Nix and MacNair (1941, 1942) as, in our opinion, they represent data of a very high order of accuracy. Moreover they have themselves used equation (2) in interpreting their data so that we can directly use their value of $\bar{\gamma}$ in equation (1). They have not given the values of $\bar{\gamma}$ explicitly, but that of another constant Q_0 which is equal to $V_0/\bar{\gamma}\kappa_0$ and from which $\bar{\gamma}$ can be obtained by substitution of the values of the other two quantities. They have not used the experimental values of C_v in their application of equation (2) but have obtained both C_v and E from an empirically derived Debye function. It is found, however, that the values of C_v obtained from this Debye function are not markedly different from the experimental values, so that the use of the Debye function does not vitiate their results. The quantity k which occurs in equation (2) has also been determined empirically by the authors. Except in the cases of molybdenum and tungsten, the values of $\bar{\gamma}$ used by them are reasonable and of the same order as those used by Grüneisen. Since k occurs in a small correction term small differences in its magnitude do not cause any large error in the evaluation of $\bar{\gamma}$ except at very high temperatures. We have therefore adopted the values of $\bar{\gamma}$ as given by these authors except in the case of molybdenum and tungsten where the values of k as used by them are extraordinarily higher than those of Grüneisen, being 6·83 and 30·63 respectively as compared with the latter's values, 2·23 and 2·28.

In these two cases it is obvious that with the use of reasonable values of k no single $\bar{\gamma}$ will explain all their data, and it must be concluded that it changes slightly with volume. We have, therefore, calculated $\bar{\gamma}$ at different volumes using Grüneisen's figures for $\frac{1}{6}(m+n+3)$. The $\bar{\gamma}$ occurring in the correction term has, however, been evaluated from the value of Q_0 of Nix and MacNair. This method has not been tried with the other metals because it was not likely to give a value of $\bar{\gamma}$ much different from that of Nix and MacNair.

In order to compare the values of p_1 obtained by the above method with Fürth's equation we make use of the equation (5) of his paper (1945)† which on substitution of his expression for the heat of sublimation Λ yields the following expression at $T=0$ for the pressure corresponding to any volume:

$$p_1V=\frac{\Lambda mn}{n-m}\left(\frac{V}{V_0}\right)^{-m/3}\left[\left(\frac{V}{V_0}\right)^{-(n-m)/3}-1\right]. \qquad \ldots\ldots(3)$$

In the above expression m and n are the exponents in the following expression for the potential energy of a pair of atoms separated by a distance r, r_0 being the equilibrium distance:

$$\phi(r)=\frac{umn}{n-m}\left[-\frac{1}{m}\left(\frac{r_0}{r}\right)^m+\frac{1}{n}\left(\frac{r_0}{r}\right)^n\right]. \qquad \ldots\ldots(4)$$

On replacing V by $V_0+\Delta V$ and using the binomial expansion we get from (3) in first approximation

$$p_1V=-\frac{\Lambda mn}{9}\frac{\Delta V}{V_0}\left(1-\frac{m+n+3}{6}\frac{\Delta V}{V_0}\right). \qquad \ldots\ldots(5)$$

This equation has been used for the calculations of p_1, using the values of Λ, m and n as given by Fürth. V_0 was extrapolated from the values of α. These values of p_1 have been compared with those obtained from Nix and MacNair's data on thermal expansion. Wherever agreement can be improved by using alternative values of m and n they are used in the calculation of p_1 from (5).

The results of the calculations are presented in the table and it will be seen that the agreement between the two sets of values is very close. The validity of Fürth's law of force (4) is thus established and it can be justifiably used in the evaluation of other thermodynamic properties of solids.

---

† We are obliged to Dr. Fürth for pointing out a printing error in equation (6) of his paper. The exponent on the right-hand side of this equation should be $-s/3$.

Substance	T	V	$\bar{\gamma}$	E (cal (g atom)$^{-1}$)	$-p_1$ (10^8 dyn cm^{-2})	
					Fürth	E/V
Aluminium	100	10·041	2·47	108	13	11
$\Lambda=67\cdot6$ kcal (g atom)$^{-1}$	200	10·092		524	53	54
$V_0=10\cdot025$ cm^3	300	10·172		1049	114	107
$m=4 \quad n=6\cdot5$	500	10·323		2180	219	219
$\theta_D=400$	800	10·586		5100	378	385
Silver	50	10·135	2·4	48	5	5
$\Lambda=67\cdot6$ kcal (g atom)$^{-1}$	200	10·218		779	86	77
$V_0=10\cdot13$ cm^3	300	10·274		1350	136	133
$m=4\cdot5 \quad n=7$	400	10·33		1940	186	189
$\theta_D=215$	500	10·393		2530	239	245
	1000	10·745		5490	492	515
Copper	150	7·12	2·14	366	42	58
$\Lambda=8\cdot17$ kcal (g atom)$^{-1}$	200	7·135		618	73	78
$m=4 \quad n=7$	400	7·207		1734	216	216
$V_0=7\cdot1$ cm^3	700	7·327		3487	431	429
$\theta_D=325$						
Gold	100	10·224	3·02	274	34	34
$\Lambda=90\cdot7$ kcal (g atom)$^{-1}$	300	10·30		1340	170	172
$m=5\cdot5 \quad n=8\cdot4$	500	10·389		2580	321	315
$V_0=10\cdot206$ cm^3						
$\theta_D=190$						
Platinum	100	9·157	2·54	231	28	26
$\Lambda=124\cdot7$ kcal (g atom)$^{-1}$	300	9·20		1330	156	154
$m=5\cdot5 \quad n=8$	400	9·225		1910	228	220
$V_0=9\cdot148$ cm^3	500	9·251		2500	302	288
$\theta_D=230$						
Lead	100	17·90	2·434	422	24	24
$\Lambda=47$ kcal (g atom)$^{-1}$	200	18·045		1000	57	57
$V_0=17\cdot795$ cm^3	300	18·2		1600	88	90
$m=3 \quad n=11\cdot5$						
$\theta_D=88$						
Molybdenum	100	9·376	1·70	113	6	8
$\Lambda=156$ kcal (g atom)$^{-1}$	300	9·40	1·78	1080	81	86
$V_0=9\cdot374$ cm^3	500	9·433	1·92	2170	186	185
$m=5 \quad n=7\cdot8$						
$\theta_D=388$						
Tungsten	100	9·578	1·524	161	15	11
$\Lambda=203\cdot9$ kcal (g atom)$^{-1}$	200	9·587	1·549	638	42	43
$V_0=9\cdot573$ cm^3	400	9·614	1·622	1760	122	124
$m=4\cdot4 \quad n=6\cdot6$	500	9·630	1·655	2340	168	169
$\theta_D=310$						

ACKNOWLEDGMENTS

Most of the numerical results presented here form the part of a thesis submitted for the degree of Doctor of Philosophy of the Banaras Hindu University by one of us (R. S. S.). We have had the advantage of having discussions on the subject with Dr. Fürth in London and take this opportunity of expressing our grateful thanks to him for his interest in the subject.

REFERENCES

BORN, M., and BRADBURN, M., 1943, *Proc. Camb. Phil. Soc.*, **39**, 104.
BRADBURN, M., 1943, *Proc. Camb. Phil. Soc.*, **39**, 113.
DAYAL, B., 1950, *J. Chem. Phys.*, **18**, 1302.
FÜRTH, R., 1945, *Proc. Roy. Soc.* A, **183**, 87.
GRÜNEISEN, E., 1926, *Handbuch der Physik*, Vol. X.
NIX, F. C., and MACNAIR, D., 1941, *Phys. Rev.*, **60**, 597; 1942, *Ibid.*, **61**, 74.

An Optically Focusing X-Ray Diffraction Camera†

By A. FRANKS‡

Birkbeck College Crystallography Laboratory, London

Communicated by W. Ehrenberg; MS. received 22nd April 1955 *and in amended form 3rd August* 1955

Abstract. An optically focusing x-ray diffraction camera is used to obtain low angle diffraction photographs. Focusing is accomplished by total external reflection either by one elastically bent optical flat to form a line focus, or by two crossed flats to form a point focus. The beam is partially monochromatized and its intensity is high. Spacings greater than 600 Å have been resolved.

§ 1. Introduction

Low angle diffraction patterns are of interest in a variety of problems, for example, in the study of long chain molecules, micellar texture of fibres, particle size determination and precipitation in alloys. Three different types of camera have been constructed to obtain these patterns.

In the first, narrow slit and pinhole systems are used to collimate the x-ray beam (which, in some cases, is monochromatized). Intensities are low and exposure times are long; nevertheless a number of authors have recorded spacings of several hundred ångström units (e.g. Bernal, Fankuchen and Riley 1938, Bear 1942, MacArthur 1943, Yudowitch 1949).

In the second type, the beam is focused and monochromatized by reflection from one or more curved crystals. For practical exposure times, Guinier (1939) finds that the largest spacing which can be resolved is approximately 150 Å, using a single monochromator. The resolution is considerably improved in double crystal monochromator cameras (Guinier and Fournet 1948, Furnas 1952, Shenfil, Danielson and DuMond 1952). Intensities, however, are low owing to the inefficiency of Bragg 'reflection'. Several difficulties are encountered with the use of monochromators, namely: large numbers of crystals must be examined in order to obtain a suitable specimen; accurate and laborious working of the crystal and crystal mounts is required; aligning techniques are difficult; resolving power is limited owing to dispersion, finite dimensions of source and scattered radiation; some crystals deteriorate after irradiation by x-rays, resulting in lowered reflectivity and an increase in scattered radiation.

In the third type of low angle camera the x-ray beam is reflected from a plane or curved polished surface. X-rays incident at a sufficiently small glancing angle are totally reflected. The critical angle of incidence is approximately proportional to the wavelength so that the reflected radiation is partially monochromatized. For CuK radiation the critical angle is approximately 11′ for light glasses and 30′ for a gold reflector. Lead glass collimators have been used by Crowfoot and Schmidt (1945) and Lely and van Ryssel (1951). Henke and DuMond (1953) describe a camera in which the beam is focused by total reflection from an annular mirror in the shape of a truncated ellipsoid. The cross section of the beam at the specimen is a ring; the apparatus has therefore only a limited use.

---

† This paper is based on a thesis which has been approved by the University of London for the degree of Ph.D.

‡ Now at the National Physical Laboratory, Teddington.

Experiments carried out by Ehrenberg (1947, 1949 a, b) led him to suggest the use of curved reflectors for the production of intense x-ray beams suitable for diffraction work. Results obtained with a low angle camera incorporating one curved reflector have briefly been described (Ehrenberg and Franks 1952). A more detailed description of this camera and also of a point focusing camera with two crossed optical elements is given below.

§ 2. A CAMERA WITH A SINGLE REFLECTOR

2.1. *Principle of the Method*

Ideally, the focusing element should be part of a cylinder of elliptical cross section. X-rays emitted from a source at one focus of the ellipse will reunite at the other focus. In practice, it suffices to bend a flat surface, the curvature being produced by the application of equal bending couples to a rectangular plate. Figure 1 shows the relative positions of source, reflector, specimen and film. The glancing angle θ (very much exaggerated) is set so that the characteristic and longer wavelength radiation only is reflected. The diffracted beams will be focused on the focal circle; owing to the depth of focus, sharp reflections will be formed at low angles on a film placed perpendicular to the beam and passing through the focus. The intensity of the incoherent radiation scattered by the optical flat, due to imperfections of the surface, Compton scattering and fluorescent radiation, is large enough to obscure the diffraction patterns. It is therefore necessary to limit the width of this parasitic radiation by placing a guard slit as close as possible to the specimen. The broken lines in figure 1 show the extreme rays of this radiation and the area surrounding the focus within which no diffraction pattern can be recorded.

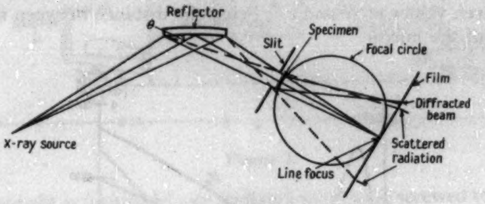


Figure 1.

2.2. *The Dependence of Intensity on the type of X-ray Source*

If the foreshortened width of the x-ray focus is small (less than 10 microns), the source itself may be at the focus of the elliptic cylinder, without introducing a slit to limit the width of the beam. It is necessary, however, to place an object slit at the focus of the elliptic cylinder when larger sources are used.

Franks (1954) has shown that the single reflector camera can be used with equal advantage either with a fine focus x-ray tube (Ehrenberg and Spear 1951) or a high intensity rotating anode tube. In both cases, the intensity per unit area of the focused beam at the film is proportional to the specific loading and height of the source. The products of these two quantities are of the same order of magnitude.

2.3. *Determination of the Optimum Camera Constants*

A compromise must be sought between the conflicting demands of high intensity and high resolution. The resolution of a low angle camera may be defined by two quantities N and N_H.

N is the first order visibility. It is equal to the largest lattice spacing which gives rise to a first order diffraction which can be detected. The diffracted beam falls just outside the central area of the film obscured by parasitically scattered radiation.

N_H is the high order resolving power and is equal to the spacing for which successive orders can just be resolved.

From a consideration of the geometry of the camera, N and N_H can be calculated as a function of its dimensions and those of the source and can be correlated with the intensity per unit area I of the focused beam at the film (Franks 1954).

The experiments were performed with a fine focus tube (foreshortened focus 6 microns wide), emitting CuK radiation ($\lambda = 1{\cdot}5$ Å), and at the critical glancing angle for glass ($\theta = 3 \times 10^{-3}$ rad). Using the numerical values, it is possible to calculate the length of the reflector, $2s$, for which I has its maximum value (i.e. $s = s_0$), for given values of N and c, c being the distance between the source and nearer edge of the mirror.

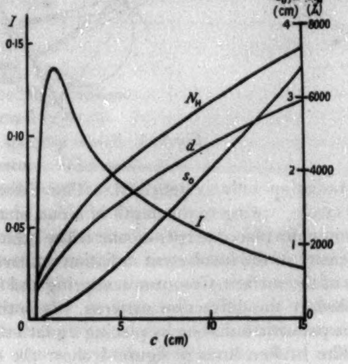


Figure 2.

Figure 2 shows representative design data obtained from the computations, in which $N = 1000$ Å. I, s_0, d are plotted against c, d being the distance from the specimen or guard slit to the film. The graph of N_H against c (figure 2) is computed on the assumption that the widths of the higher orders are equal to that of the source. The graph shows that the intensity increases to a maximum value as c decreases, i.e. as the camera becomes shorter, but after a further decrease in c the intensity rapidly falls to zero. This is due to the fact that the width of the focus is finite; as c decreases the width of scattered radiation decreases while simultaneously s_0 decreases. At a finite distance from the source, the width of the scattered radiation equals the width of the focus, but at this point s_0 vanishes, and consequently the intensity is zero. At the value of c where I is a maximum the distance *d* is less than 1 cm. Where precision measurements of spacings are required it is desirable to increase this distance. This requirement is effected at the expense of intensity.

2.4. *Description of the Focusing Camera*

The effective part of the apparatus is shown in figure 3. In constructing it, use was made of existing apparatus. The reflector is a glass optical flat 4 cm long, 1·5 cm wide and 4·5 mm thick. It is supported on one side by two 'silver steel' rods 1·27 mm in diameter, placed 5 mm from each end. The rods are undercut in the middle to allow the x-rays to pass. On the other side a variable bending couple is applied by means of a micrometer screw which bears against the steel spring, which in turn presses against the edges of the reflector. The rods are loosely held in a support (not shown), against an accurately machined reference surface, a steel bar of cross section 1 inch square.

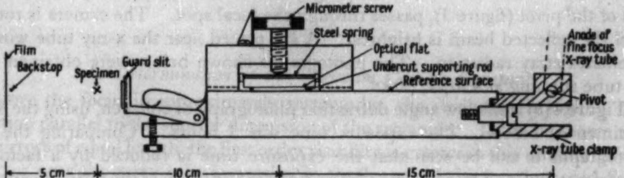


Figure 3.

The guard slit is formed by two Pt flats $\frac{1}{2}$ mm thick, screwed to a brass frame, and is set, with a feeler gauge, to a width slightly exceeding the beam width. The backstop, a thin tungsten wire (0·3 mm in diameter), is placed immediately in front of the film, and is mounted eccentrically on a vertical spindle. By rotation of the spindle, the wire is brought into the path of the beam, thus preventing the film from being fogged during long exposures. The x-ray focus-to-film distance is 30 cm. (Once the main dimensions (d, L, s), $2L$ being the distance from source to focus, have been specified, the width of the guard slit and diameter of backstop can be obtained very conveniently from a scale drawing, actual scale along the length of the camera and with a magnification of 100 at right angles.)

2.5. *Experimental Technique and Results*

A 5 kw rotating anode tube of the type designed by Clay (1934) was employed in the preliminary experiments. Since the foreshortened width of the focus is approximately 1 mm, it was necessary to place an object slit (10 microns wide) near the tube. By using a feeler gauge in contact with the reference surface (figure 3), the slit was set so that the x-ray beam was incident on the reflecting surface at an angle of 3×10^{-3} rad.

The grain of the conventional fluorescent screen prevents any change in focal line width from being observed. The camera is focused visually by using a single transparent crystal of CsI–Tl (Franks 1953), 0·2 mm thick, as fluorescent screen. The fluorescent spot is observed through a microscope (400 ×). Alternatively, the reflector can be focused by taking a series of trial photographs for different settings of the micrometer focusing screw. A typical photograph obtained with the camera and rotating anode tube is shown in figure 4 (*a*) (Plate). It is the low angle diffraction pattern of collagen from a kangaroo tail tendon. The exposure time was 13hours, at 70 mA, 33 kvp; the second order (320 Å) is resolved. The photograph is an enlargement, the magnification being indicated by the scale. This photograph and all the subsequent ones were taken with CuK radiation on Ilford industrial G film. The emulsion on one side of the film is removed after development since the patterns do not overlap when viewed at right angles to the plane of the film.

All the other diffraction photographs shown below were obtained with a fine focus tube (Ehrenberg and Spear 1951), the x-ray source being placed at the focus of the reflection ellipse. The dimensions of the foreshortened focal spot are approximately 40×6 microns. The camera and tube now form one unit and it is important to prevent any relative motion between these two components. The camera is clamped to the tube and can be rotated about the x-ray focus using a slightly modified version of the clamping system described by Jeffery (1952). The axis of the pivot (figure 3), passes through the focal spot. The camera is rotated until the reflected beam is brightest. A slit placed near the x-ray tube window eliminates stray radiation. The photographs shown below were obtained with the tube running at 450 μA, 50 kv.

Figure 4 (*b*) is the low angle diffraction photograph of collagen, using the same specimen as before. The exposure time was 3 hours. Comparing the two photographs it can be seen that the exposure time is reduced by a factor of approximately 5. This is in fair agreement with the value of the ratio of exposure times obtained theoretically. Figure 4 (*c*) is the low angle photograph of a specimen of tobacco mosaic virus gel prepared according to the method described by Bernal and Fankuchen (1941). The innermost reflection corresponds to a spacing of 132 Å. The x-ray beam is perpendicular to the particle axis. The exposure time was 3 hours.

2.6. *Discussion of Experimental Results*

The calculated first order visibility of the camera is just over 700 Å. Films exposed for an hour or longer (e.g. figure 4 (*b*)) show some fogging round the shadow of the backstop. This fogging is not directly due to the relatively intense incoherent radiation scattered by the reflector, but to the parasitic radiation scattered by the guard slit. The width of the incoherent radiation is sharply defined as can be seen from figure 4 (*d*), a 15 minute exposure showing the diffraction photograph of collagen from a rat tail stained with phosphotungstic acid, the backstop being 0·125 mm diameter instead of 0·3 mm; the 640 Å spacing is now resolved. The first order visibility for this short exposure agreed more closely with the computed value. The higher order resolution agrees with the calculated value and is of the order of 10 000 Å.

§ 3. A CAMERA WITH TWO REFLECTORS

The camera described above gives a one-dimensional diffraction pattern only. Much more information would be contained in a two-dimensional pattern. This can be realized by introducing a second mirror perpendicular to the first, so that the pair of mirrors produces on the film a one to one image of the foreshortened x-ray focus 6×40 microns in size. Preliminary calculations showed also that the second mirror would bring about a worthwhile increase in intensity. The arrangement of the apparatus is shown in figure 5 and resembles the x-ray microscope of Kirkpatrick and Baez (1948). The usual formulae relating magnification, object and image distances are valid for x-rays reflected at grazing incidence. The aberrations are least at unit magnification (Baez 1948). In the double focusing camera, aberrations are kept as small as possible by having both mirrors close to the centre of the system.

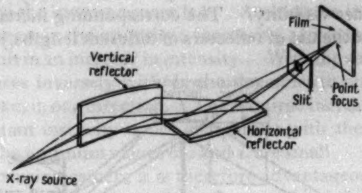


Figure 5.

3.1. *Determination of the Optimum Camera Constants*

Since the focus is approximately rectangular, the resolution is highest in a direction parallel to the shorter dimension of the rectangle. It can be shown that for mirrors of equal length, the first order visibility is greater in this direction if the reflector near the x-ray tube is parallel to the length of the foreshortened focus (vertical). Results obtained with this particular arrangement are discussed below. The intensity is independent of the order in which the mirrors are placed.

Figure 6 is a plan view, showing the geometry of the camera; the mirrors are of equal length.

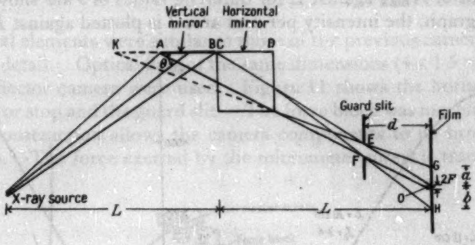


Figure 6.

AB is the reflecting surface of the vertical mirror, length $2s$. CD is the horizontal mirror, length $2s$. EF is the width of the guard slit. O is the centre of the focused beam. $2L$ is the distance between source and focus. d is the slit (or specimen) to film distance. OG is the width of scattered radiation on one side of O, and is due to the horizontal mirror.

The dashed symbols have the equivalent significance in the elevation (figure 7). The radiation O′G′ is scattered by the horizontal mirror only. A stop, the mirror stop, bisects C′D′ and absorbs very nearly all the radiation scattered by the vertical mirror.

From the geometry of figures 6 and 7, the horizontal and vertical first order visibility (N_h and N_v respectively) can be calculated (Franks 1954). The variation of N_h and N_v with d, for various values of L and for $s=2$ cm is shown in figure 8. The intensity per unit area at the focus is proportional to the product of the angular apertures of both reflectors. The values of the intensity shown in figure 8 are in arbitrary units and were computed for $s=2$ cm. (It can be seen from figure 6 that the length of the first mirror AB may be increased by a length AK without affecting the first order visibility. The corresponding increase in intensity does not normally warrant the use of reflectors of different lengths.)

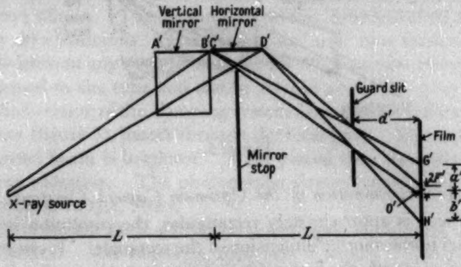


Figure 7.

The shapes of the curves of figure 8 do not vary much when plotted for widely differing mirror lengths and specimen-to-film distances. $N_{\max}$ is the largest first order spacing which can be observed and is given by the peaks of the 'h' curves.

The graphs of $N_{\max}$ against L for different values of s are shown in figure 9. On the same graph, the intensity per unit area I is plotted against L for the same

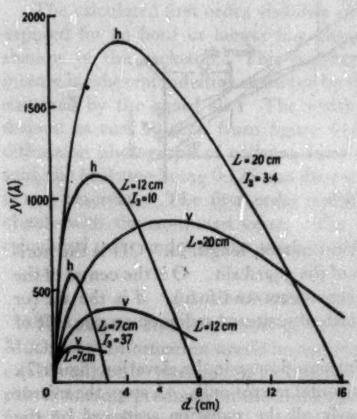


Figure 8.

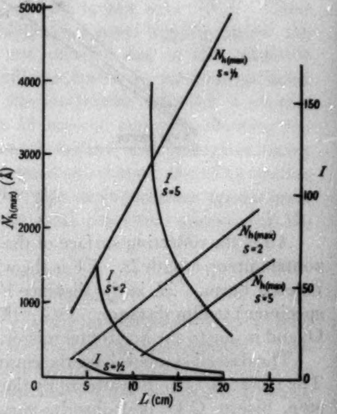


Figure 9.

values of s. The curves show that for a given first order visibility over the range calculated, the greatest intensity is obtained by using reflectors as large as possible with the corresponding optimum values of L and d. In general, increasing the mirror length increases the intensity but decreases the first order visibility. On the other hand, increasing the distance between source and focus decreases the intensity but increases the visibility. The intensity per unit area at the focus depends on the product of the apertures of the mirrors and does not decrease with distance according to the inverse square law. An increase in L therefore causes a relatively small decrease in the intensity; increasing s to reduce the visibility to its initial value will result in an increase in intensity. With the single reflector camera, the intensity increases inversely with L, the dependence on L being due to the divergence of the beam in one direction. The required increase of s with increasing L to maintain constant intensity is not compatible with the first order visibility conditions, so that the optimum values of s and L are small.

In the double focusing camera it is therefore advantageous to work with large values of s and L for maximum intensity. Large values of s and L are associated with correspondingly large values of the optimum specimen-to-film distance; this is very convenient experimentally and also allows measurements of the diffraction patterns to be made with greater accuracy.

3.2. *Description of the Double Focusing Camera*

The components of the camera were mounted on an optical bench (The Precision Tool and Instrument Co. Ltd.), which was supplied with bench saddles having lateral and vertical screw controlled traverses. A photograph of the camera is shown in figure 10 (Plate). The camera is clamped to a fine focus tube and stands on a concrete table top, 3 in. thick. The concrete slab rests on anti-vibration mounts screwed to an angle iron table. This massive construction is necessary to suppress the excessive vibrations from which the laboratory suffers.

The optical elements were similar to those of the previous camera, but differed somewhat in detail. Optical flats of the same dimensions (4 × 1·5 × 0·45 cm) as in the single reflector camera were used. Figure 11 shows the horizontal reflector unit, the mirror stop and the guard slit. The force block was machined from brass rod. This construction allows the camera components to be brought closer to the reflectors. The force exerted by the micrometer screw is transmitted to the

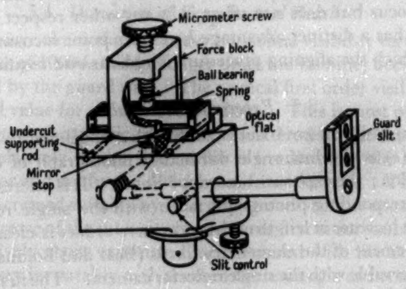


Figure 11.

spring via a 1/16 in. ball bearing, to prevent any lengthwise movement of the reflector on rotation of the screw.

A spring loaded movement of the guard slit is provided by two screws acting against the force set up by distortion of the thin part of the supporting rod.

The vertical reflector unit is similar to the horizontal one and consists of a force block screwed to a vertical plate, which is clamped to a bench saddle.

The film holder is shown in figure 12. The backstop is a bent tungsten wire attached to a vertical spindle, which is mounted on the film holder frame. The wire can be taken out of the path of the beam by rotating the lever fixed at right angles to the spindle and returned to its original position by use of the stop. This is a useful device for obtaining the trace of the main beam. The backstop is correctly set in the path of the beam by means of the two traverses of the film holder saddle.

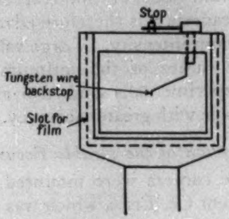


Figure 12.

The focus-to-film distance was made equal to 24 cm. It can be seen from figure 8 that with 4 cm mirrors the maximum theoretical first order visibility is 1100 Å and the first order vertical visibility for the same specimen-to-film distance is 400 Å. Aligning of the camera components is greatly facilitated by use of the single crystal fluorescent screen and after some experience has been gained, the whole procedure takes no longer than half an hour. The method used, was to set the reflecting surface approximately parallel to the direction of the beam, then to move the reflector laterally until the reflected beam flashed into view, the glancing angle of incidence then being slightly less than the critical angle. The reflectors were then bent until the beam was focused. It is important to realize that if this procedure is followed, changing the orientation of the reflector merely changes the position of the focus but does not affect it in any other respect. The optically focusing camera has a distinct advantage here over point focusing crystal monochromators in which the aligning procedure is tedious and requires great care.

3.3. *Experimental Results*

The photographs in figure 4 were taken with a specimen-to-film distance of 3 cm. Figure 4 (*e*) is the low angle diffraction photograph of collagen from a kangaroo tail tendon; the exposure time was 3 hours. It is interesting to compare this with the corresponding photograph taken with the single reflector camera, figure 4 (*b*). The increase in length of the successive orders is clearly visible; it is due to the arrangement of the elementary fibrils (Bear and Bolduan 1950). This effect was not observable with the single reflector camera. The first order (640 Å) is now resolved and the lower orders are very much overexposed, showing that the intensity per unit area is greater than before. The intensity per unit area is increased by a factor of about 7 which is in fair agreement with the factor 10 deduced theoretically. The increased lengths of the higher order reflections reduces their relative intensities. Figure 4 (*f*) is the low angle diffraction pattern of rat tail stained with phosphotungstic acid, exposure time 10 minutes. There is no fogging round the shadow of the backstop. With these short exposures, spacings of over 1000 Å have been resolved. The behaviour of the camera parallel to the length of the line focus is demonstrated in figure 4 (*g*). The stained rat tail is vertical, the exposure being 40 minutes. The highest vertical spacing which can be resolved is about 500 Å. The higher order resolving power is reduced considerably. Figure 4 (*h*) is the low angle pattern of heat relaxed terylene (fibre horizontal); exposure time is 30 minutes. Figure 4 (*i*) is the diffuse low angle scattering pattern of a single crystal of an Al–Cu alloy (3% Cu), containing groups of Cu atoms separated out from the supersaturated solution (Guinier 1952). The exposure time was 3 hours.

Figure 4 (*j*) is the low angle diffraction pattern of tobacco mosaic virus, showing the hexagonal packing of the virus rods, exposure time 3 hours. Figure 4 (*k*) is the low angle diffraction pattern of a gliadin film, exposure time 12 hours. The film was obtained by evaporation of a gliadin solution from Atle wheat. The normal to the film was perpendicular to the x-ray beam and at about 45° to the horizontal. The arcs superimposed on the 50 Å ring clearly indicate orientation parallel to the normal. Figure 4 (*l*) is the diffraction photograph of decayed Manitoba gluten, exposure time 12 hours. The boundary of the intense low angle scattering corresponds to a spacing of 220 Å. The sharp cut-off of the diffuse scattering coincides with an intense 50 Å ring. The diffraction effects from gliadin and Manitoba gluten have not been published before. Both specimens were kindly prepared by Dr. W. Traub of this College.

Figure 4 (*m*) is the low angle scattering pattern of coal from Skelton Grange, exposure time 1 hour.

The use of the focusing camera for back reflection work has not yet been seriously investigated. Preliminary experiments however have yielded some promising results. Figure 4 (*n*) shows a back reflection photograph of Al ($4\frac{1}{2}$ grains per mm), which had been creep tested for 696 hours at 200°c, under a stress of $\frac{1}{3}$ ton in^{-2}; the exposure time was 2 hours. The specimen was kindly supplied by Dr. D. McLean of the National Physical Laboratory.

3.4. *Discussion of Experimental Results*

For short exposure times the first order horizontal visibility very nearly equals the theoretical value of 1000 Å, but for longer times the visibility decreases owing to radiation scattered by the guard slit. The vertical first order visibility is greater than the calculated value for short exposure times. This is most probably due to the fact that the vertical parasitic radiation is scattered by the horizontal mirror only and is thus relatively weak; this mirror is irradiated by the beam reflected from the first mirror and not by the direct beam from the tube. The higher order visibility may be increased suitably by enlarging the scale of the apparatus, and this also leads to an increase of the intensity for the same first order visibility.

The comparatively short exposure times required with the double focusing camera compare very favourably with those of the low angle cameras mentioned in the introduction, for photographs of comparable resolution.

ACKNOWLEDGMENTS

I desire to express my gratitude to Professor J. D. Bernal, for the interest he has taken in this work. I am deeply indebted to Dr. W. Ehrenberg, who initiated this research, for continued advice and many helpful and encouraging suggestions.

REFERENCES

BAEZ, A. V., 1948, *X-ray Optical Images* (Stanford : University Press).
BEAR, R. S., 1942, *J. Amer. Chem. Soc.*, **64,** 727.
BEAR, R. S., and BOLDUAN, O. E. A., 1950, *Acta Cryst., Camb.*, **3,** 236.
BERNAL, J. D., and FANKUCHEN, I., 1941, *J. Gen. Physiol.*, **25,** 111.
BERNAL, J. D., FANKUCHEN, I., and RILEY, D. P., 1938, *Nature, Lond.*, **142,** 1075.
CLAY, R. E., 1934, *Proc. Phys. Soc.*, **46,** 703.
CROWFOOT, D., and SCHMIDT, G. M. J., 1945, *Nature, Lond.*, **155,** 504.
EHRENBERG, W., 1947, *Nature, Lond.*, **160,** 330; 1949 a, *J. Opt. Soc. Amer.*, **39,** 741; 1949 b, *Ibid.*, **39,** 746.
EHRENBERG, W., and FRANKS, A., 1952, *Nature, Lond.*, **170,** 1076.
EHRENBERG, W., and SPEAR, W. E., 1951, *Proc. Phys. Soc.* B, **64,** 67.
FRANKS, A., 1954, *Ph.D. Thesis*, University of London.
FRANKS, J., 1953, *Brit. J. Appl. Phys.*, **4,** 377.
FURNAS, T. C., 1952, *Thesis*, Massachusetts Institute of Technology.
GUINIER, A., 1939, *Ann. Phys., Paris*, **12,** 161; 1952, *X-ray Crystallographic Technology* (London : Hilger and Watts).
GUINIER, A., and FOURNET, G., 1948, *C.R. Acad. Sci., Paris*, **226,** 656.
HENKE, B., and DUMOND, J. W. M., 1953, *Phys. Rev.*, **89,** 1300.
JEFFERY, J. W., 1952, *J. Sci. Instrum.*, **29,** 385.
KIRKPATRICK, P., and BAEZ, A. V., 1948, *J. Opt. Soc. Amer.*, **38,** 766.
LELY, J. A., and VAN RYSSEL, T. W., 1951, *Philips Tech. Rev.*, **13,** 96.
MACARTHUR, I., 1943, *Nature, Lond.*, **152,** 38.
SHENFIL, L., DANIELSON, W. E., and DUMOND, J. W. M., 1952, *J. Appl. Phys.*, **23,** 854.
YUDOWITCH, K. L., 1949, *J. Appl. Phys.*, **20,** 174.

The Measurement of Comatic Aberrations by Interferometry

By J. W. GATES
Light Division, National Physical Laboratory, Teddington, Middlesex

MS. received 25th March 1955 *and in amended form 28th July* 1955

Abstract. The path differences from point to point between an optical wave front and its mirror image are antisymmetrical with respect to the common line about which the inversion of the mirror image has occurred. Features of the wave front which depend only on even powers of the distance from the common line disappear, while features which depend on odd powers of the distance are doubled. Thus by suitable choice of the common line the comatic errors and the distortion in the wave front may be shown up independently of the effects of change of focus, spherical aberration, astigmatism and field curvature. The analysis may be carried out on interferograms made by comparison with a known reference wave front, or by the use of the interferometer described, which compares the wave front with its own mirror image directly. This interferometer may also be used as a shearing interferometer.

---

§ 1. Introduction

The problem of analysing interferograms is notoriously difficult when many aberrations are present simultaneously. On the other hand, the patterns produced by a wave front in which only one or two aberrations are present can easily be interpreted. Some of the first investigations in this direction were carried out by Kingslake (1926), who produced the interferograms on a Twyman interferometer. In this instrument the wave front under investigation is compared with a reference wave front of known shape; in order to simplify the analysis, the shape of the reference surface should be a perfect sphere. To produce a reference wave front of the quality required demands the highest standard of workmanship on the optical parts of the interferometer, some of which, at least, must be as large as the system to be examined. The need for a reference wave front and the large, accurate optical parts to produce it was avoided in the shearing interferometer devised by Bates (1946, 1947). Not only is this interferometer very compact and convenient for the examination of large optical systems, but the shearing principle makes possible a very great simplification of the process of analysis of the interferograms. More recent developments in the design and use of shearing interferometers have been described by Drew (1951) and by Brown (1954). The latter gives some examples of the calculation of the primary aberrations. Shearing interferometers dispense with the comparison wave front by providing two sights of the system under test, so that the wave front produced by the system may be compared with itself. Any point on one wave front can be made to coincide with a point on any zone of the other wave front by altering the shear between the two wave fronts. In some cases a complete analysis of the shape of the wave front can be made by this procedure.

The present paper is concerned with the development of the principle of comparing two parts of the same wave front, or two parts of the same map of a wave front, in order to isolate and measure the comatic contributions to the wave front aberration.

§ 2. Comatic Aberration

The presence of coma or the related aberrations in an optical system is often objectionable because of the unsymmetrical nature of the image patch and because the 'centre of gravity' of the patch may be displaced when the aperture of the system is changed. Other aberrations of a symmetrical nature may be more tolerable, although under some conditions (e.g. vignetting of the pupil) an image shift may be produced by a change of aperture when only symmetrical aberrations are present. It is often useful, in any case, to be able to measure small amounts of comatic aberration in the presence of larger amounts of symmetrical aberration. The asymmetry which distinguishes the comatic wave front aberrations is well brought out in the classification of the aberrations given by Hopkins (1950). This system employs the coordinates indicated in figure 1. The optical system is shown dotted; the optical axis is OO′. PP′ is an oblique principal ray which cuts the pupil ABCD of the system in S. The obliquity of this ray is measured by the intercept O′P′ in the focal plane, and is expressed as a fraction σ of the maximum value of O′P′. Q is a point in the pupil and has the coordinates r (expressed as a fraction of the maximum semi-aperture) and ϕ, the angle SQ makes with the tangential meridian SB. Hopkins represents the total wave front aberration by the sum of a series of terms, each of which is the product of a coefficient and powers of r, σ and $r\cos\phi$. The comatic aberrations are those which involve odd powers of $\cos\phi$. An investigation of the variation of the wave aberration over the pupil for a number of values of σ should make possible the identification of the aberration terms present and the assignment of a value to each coefficient. When all the aberrations are present the task is

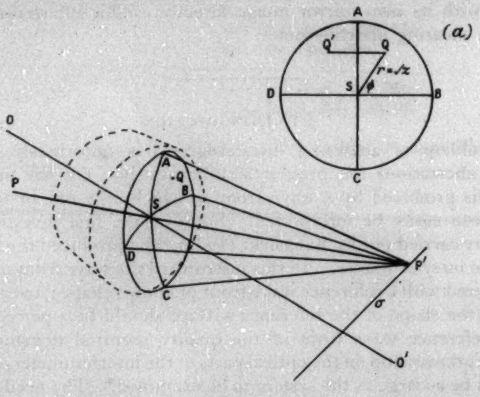


Figure 1. Coordinate system for analysis of wave front aberration.

rather difficult, since there are five coefficients to be determined simultaneously for primary aberration only, and fourteen for primary and secondary together. By eliminating all the contributions to the aberration which depend on even powers of $\cos\phi$, the number of coefficients to be determined is reduced to two for the primary aberration only, or six when both primary and secondary aberrations are present. When the purpose of the investigation is to examine image quality only, these numbers are each reduced by one, since one term in each order specifies a distortion, which has no effect on the structure of the image of a small object. A further simplification is introduced in this paper by considering only a single value of the obliquity σ. Each aberration term then takes the same form as one of the terms in each higher order, since these terms differ only in the power of the σ factor. For example, spherical aberration terms of the form r^4, $\sigma^2 r^4$, $\sigma^4 r^4, \ldots$ occur. By fixing σ, each of these terms becomes a function of r^4 only, and the whole series may be summed and treated as a single term. The number of coefficients to be determined is thereby reduced still further. If we now write $r\cos\phi = x$ where x is the perpendicular distance of Q from the sagittal meridian AC of the pupil in figure 1 (*a*), the wave aberration at Q may be expressed as a sum of terms of the form $a_{mn}x^m r^{2n}$ where each a is a constant for the optical system, and m and n are integers. Further, we can write $r^2 = z$ and the wave aberration at Q may then be expressed as

$$w_{\text{total}} = \sum a_{mn} x^m z^n. \qquad \ldots\ldots(1)$$

The first few terms of this series for which m is odd are

$$w_{\text{comatic}} = a_{10}x + a_{11}xz + a_{12}xz^2 + a_{30}x^3 + a_{13}xz^3 + a_{31}x^3z + \ldots \qquad \ldots\ldots(2)$$

In this series, the first term includes all the distortion contributions of all orders. The second term is sufficient to express the primary coma contribution, though the term will include higher orders if these are present. Similarly, the first four terms taken together are sufficient to express all the comatic and distortion contributions in the primary and secondary orders, and the six terms quoted are sufficient for the first three orders.

In order to analyse the comatic aberrations into the contributions set out above, it is necessary to measure w_{comatic} for a number of points in the pupil, and to solve the series of simultaneous equations obtained to find the constants $a_{10}, a_{11}, \ldots$. This was done in two stages. Equation (2) becomes, after dividing through by x and rearranging (and dropping the subscript from w_{comatic}),

$$w/x = (a_{10} + a_{11}z + a_{12}z^2 + a_{13}z^3 + \ldots) + x^2(a_{30} + a_{31}z + \ldots) + \ldots \qquad \ldots\ldots(3)$$

In this expression w/x is expressed as a polynomial in x^2 in which each coefficient of x^{2p} (where p is any integer) is a polynomial in z. For the example to be given, the two terms above were sufficient. Equation (3) was treated as a linear equation in w/x and x^2, for a series of constant values of z, giving values of the polynomials in z for each value of z. These two sets of simultaneous equations were then solved to obtain the coefficients $a_{10}, a_{11}, \ldots$, taking each polynomial as far as was justified by the measurements of w from the interferograms. If a complete evaluation of all the aberration coefficients of each order was required, it could be carried out by obtaining the coefficients $a_{10}, a_{11}, \ldots$, as above for each obliquity σ, and expressing each coefficient as a polynomial in σ. A series of simultaneous equations could thus be formed which permit a value for each of Hopkins' aberration coefficients to be obtained.

§ 3. Measurement of Comatic Aberration

In figure 1 (*a*), the point Q in the pupil has the coordinates x, $\sqrt{z}$. Q′, the mirror image of Q in the sagittal meridian, has the coordinates $-x$, $\sqrt{z}$. Consequently all the terms in w_{total} in which m is even will have the same value at Q and at Q′, while terms for which m is odd will have equal magnitude but opposite signs. Thus if we are able to measure the total wave front aberration at two points symmetrically situated with respect to the sagittal meridian, the difference of the two will be equal to twice the sum of the individual comatic contributions, and all other contributions will cancel:

$$w_{\text{total}} - w'_{\text{total}} = 2w_{\text{comatic}}. \qquad \ldots\ldots(4)$$

Measurements of the total aberration may be made from interferograms which are obtained by comparing the unknown wave front with a perfect reference wave front, as in the Twyman interferometers described by Candler (1951). Examples of the interferograms are given in figures 2 and 3 (Plate I), which were obtained with a camera lens of 20 inches focal length set at 10° off axis on the Twyman camera lens interferometer. In figure 2 the reference surface has been set parallel to the mean plane of the wave front. In figure 3 a tilt of the reference surface has been made to cause the bands to close up and to produce a pattern in which the order of interference increases progressively all the way across. Both interferograms are accurate maps of the unknown wave front, but the second is more easily measured, since a possible uncertainty in the order of some bands is avoided, and the points measured may be spread more evenly over the surface. It will be seen that the pattern is nearly symmetrical, indicating that most of the aberration consists of even-order terms.

The measurement and comparison of the two sides of the interferogram may be combined if two copies of the pattern are prepared and measured in some form of comparison microscope. A microscope which had been specially constructed for another kind of related measurement, and was equipped with a pantograph for plotting directly the measurements as they were made, was adapted for the purpose. This instrument has been described elsewhere by Gates (1954).

§ 4. Separation of Coma Pattern by Moiré Fringes

A quick way of separating the comatic parts of the wave front aberration can be employed if two transparent photographic copies of the Twyman interferogram are available. The interferogram should have nearly parallel bands crossing the pupil in the tangential direction, as in figure 3. The two prints are put together face to face with the two sagittal meridians in coincidence, when the moiré pattern shown in figure 4 (Plate I) is obtained. The pattern is symmetrical in appearance about a vertical axis, though strictly the path differences represented by the pattern are of opposite signs on the two sides. The central band is one of maximum contrast, because the spacing of the bands in the two prints coincides. On either side there are successive bands of minimum contrast formed of points where a dark band of one print lies on a clear band of the other. Thus each band is a locus along which there is an odd number of half-bands difference between the two sides. Because light passes twice through the system under test, each band represents the locus in the wave front, formed after a single traverse of the system, along which the comatic aberration is constant and has a value of an odd multiple of $\frac{1}{8}$ wavelength, referred to the sagittal meridian. This method, unlike the others described here, is not suitable for the measurement of comatic aberration much less than half a wavelength.

§ 5. The Double Image Interferometer

In most two-beam interferometers, an original wave front is divided into two parts which are later recombined after one part has passed through the system to be tested. Under suitable conditions, interference will show up the aberrations and other defects in the system. The interferometer to be described here is so arranged that one part of the wave front is reversed, and both parts are then passed through the system to be tested along similar paths. A second reversal of one of the two parts of the original wave front brings the two into coincidence so that interference can occur, but because of the reversal, only the asymmetrical variations of path in the optical system are detected.

The interferometer is automatically compensated for use with white light, since the central rays of the two parts of the original light pencil follow exactly the same path but in opposite directions. Theoretically an extended source could be used, but in practice small errors in construction restrict the illuminating aperture which may be used. A note on this restriction and its estimation is given in the appendix.

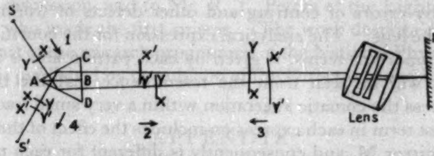


Figure 5. The double-image interferometer applied to the measurement of coma.

One form of the arrangement is shown in figure 5, applied to the testing of a photographic lens. ABC and ABD are two exactly similar right angled prisms, each with one 60° angle, made by cutting a single prism in two. The face AB of one of the prisms is coated with a semi-reflecting film and the two prisms are cemented at the faces AB in such a way that the faces CB and BD are coplanar within a very small limit. A small source is placed as shown, at S, so that light enters the prism through the face AC and is divided at the interface AB into two beams of equal amplitude. The two beams are totally reflected at AC and AD respectively, and emerge through the face CD on paths which converge slightly towards the system under test. A mirror M is set behind the system under test so as to focus a reflected image of the source S back on itself. The mirror may be a sphere of any curvature set with its centre coincident with the second principal focus of the system. In testing lenses intended to be used with one conjugate infinite, the mirror must be plane and the system must be turned with the front towards the mirror. This is the case which is illustrated in this paper. In the figure the prism is enlarged out of proportion for the sake of clarity; in practice the angle at which the two halves of the light beam converge is very much smaller. Successive positions of the parts of the original wave front XY are shown. The wave front at 1 is split in the prism into two parts XY and X'Y', which emerge at

2 as mirror images of each other. After passing through the lens and being reflected at the mirror M, the two wave fronts again pass through the lens and emerge at 3. Here the two parts of the wave front are diverging, and each retraces in the opposite sense the path originally traced by the other. The two parts re-enter the prism, where each is further divided into two parts. One part of each returns towards S and is lost; the remaining parts of the two wave fronts leave the prism at 4 and are congruent except for unsymmetrical components of path difference introduced by the lens. An eye looking into the interferometer prism from S′ sees two images of the pupil of the system, which are mirror images and which may be made to overlap by rotating the prism about an axis perpendicular to the plane of the figure. Rotation of the mirror M, also about an axis perpendicular to the plane of the figure, produces opposite tilts of the two reflected wave fronts at the position 4. This provides a control of the separation of the interference bands which appear to cross the image of the pupil of the system. The pattern of bands is similar to that which would be formed by combining two copies of a Twyman interferogram to form a moiré pattern, as described above; a series of such patterns, obtained with the same lens at the same obliquity as in figures 2, 3 and 4, is shown in figure 6 (Plate II). The mirror M is set at progressively increasing angles through the series. Each of these patterns has been analysed as described above, combining measurements for each of the four similar quadrants of each pattern, in order to reduce the effect of the distortions in the pattern caused by errors of centring and other defects of workmanship in the construction of the lens. The analytical expression for the comatic errors, giving the three most important terms, is given by each pattern and is also plotted for comparison. It will be seen from the resemblance that the three terms are sufficient to express the comatic aberration within a very small fraction of a wavelength. The first term in each expression includes the effect of the tilt introduced in rotating the mirror M, and consequently is different for each pattern, but the other terms represent comatic aberration terms which depend only on the construction of the lens and the obliquity angle, and are therefore the same for all the patterns.

§ 6. SHEARING INTERFEROGRAMS

As already described, the two images of the pupil of the system may be sheared by rotating the prism, and at any state of shear the separation of the interference bands may be controlled by tilting the mirror M. Some examples of the patterns obtained are shown in figure 7 (Plate I). When the aberrations to be measured have zonal symmetry, or are symmetrical about the sagittal meridian, no change is produced as a result of the inversion of one of the images of the pupil, and the patterns which are produced are exactly similar to those which can be obtained from the types of shearing interferometers described by Bates, Drew and Brown. The prism is more compact than any of these interferometers, though it has the disadvantage that it is not possible to introduce a tilt of the wave fronts about an axis in the plane of the diagram in figure 5.

It seems that prisms of this kind were first applied by Kösters (1931) who points out that any combination of two exactly similar prisms can be used as an interferometer in the way described as long as one of the angles of the prism adjoining the common face is exactly half the angle which is opposite the common face. Thus one type of prism which could be employed is the beam splitting cube which is already in use for other purposes. This consists of two right-angled isosceles prisms joined along their hypotenuse faces.

§ 7. CONCLUSION

This paper shows that a very great simplification of the problem of analysing a complex interferogram of an optical system can be made by separating the odd and even parts of the wave aberration. One interferogram is sufficient to determine the total comatic contribution at each point in the pupil, and it has been shown how the aberration can be broken down into its constituent terms, with good agreement between the original wave front and the map obtained by plotting the sum of the major terms. The principle can be extended to deal with other types of aberration in a similar way; for example, the even order aberrations could be separated out from the complex interferogram, and intercomparison of two such maps sheared through 90° about the centre of the pupil will make possible the measurement of astigmatic contributions. Further investigations of this kind are planned.

ACKNOWLEDGMENTS

Acknowledgment is made to Dr. H. Barrell for drawing attention to the Kösters patent, and for discussion, and to Mr. R. T. Peters of the Light Division Glass-working Shop, who made the first prisms. The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

APPENDIX

Limitations to the Size of Source which may be used with the Double Image Interferometer

As suggested in the paper, it is convenient to make the two parts of the prism by cutting a single prism in two. The two halves of the completed prism will then be symmetrical, except for the cement layer which adds to the optical thickness of the uncoated half prism. In addition to the small path difference introduced by this layer, there may be small misalignments of the two halves, and the layer may be wedged. The path difference is automatically compensated in the arrangement described, but the misalignments and difference of angles set a limit to the size of source which may be used. Fortunately, the size and shape of the source which may be used is easily found in practice.

Figure 8 shows a section of the prism and represents the surfaces encountered by each of the two possible ray paths in figure 5. A wave front passing through the system from the region of P is reflected and returned towards P′. The transmission through the system under test and the reflection is to be regarded as limited to a small region Q of the pupil of the system over which the phase variations introduced by the system are small. In these circumstances the two parts of the original wave front will coincide except for a small relative tilt introduced by differences in the angle of the two half prisms. The tilt is of the magnitude $2(n-1)(\gamma-\alpha+\beta-\delta)$ where n is the refractive index of the material of the prisms, and α, β, γ, δ are the angles shown.

Similarly a section at right angles to the one shown gives an expression for the relative tilt of the two wave fronts in terms of the pyramidal errors in the prisms, but the expression is not simple except for special cases. The variation of brightness seen in a region conjugate with Q (if Q is now regarded as moving across the whole pupil) will not be appreciable unless the two wave fronts arriving at P′ from each area Q are in phase over the whole region of P′. If there are angular errors, good contrast in the interferogram will only be obtained by restricting P or P′. In practice this can easily be checked, since an extended source can be placed at P, and the relative tilt in the two wave fronts is shown by the interference bands seen in the region of P′. The largest aperture which can be employed is a slit of a width less than one band, oriented similarly to the bands in the pattern, and placed so as to coincide with the achromatic band in the case of white light testing. In the prisms actually used, it was found convenient to employ a pinhole rather than a slit, and this was usually about 0·5 mm in diameter.

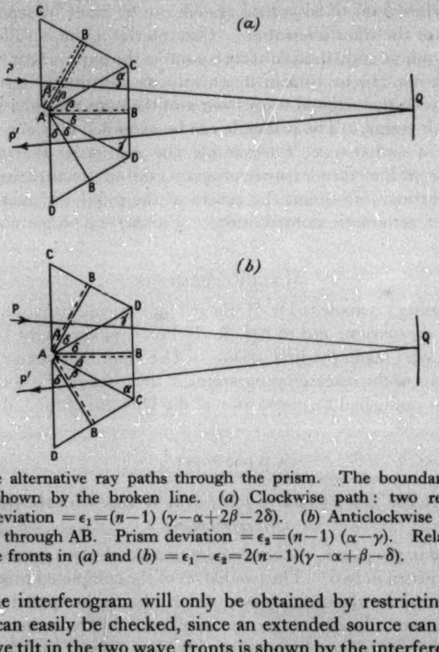


Figure 8. The alternative ray paths through the prism. The boundary of the cement layer is shown by the broken line. (*a*) Clockwise path: two reflections at AB. Prism deviation $=\epsilon_1=(n-1)(\gamma-\alpha+2\beta-2\delta)$. (*b*) Anticlockwise path: two transmissions through AB. Prism deviation $=\epsilon_2=(n-1)(\alpha-\gamma)$. Relative tilt between the wave fronts in (*a*) and (*b*) $=\epsilon_1-\epsilon_2=2(n-1)(\gamma-\alpha+\beta-\delta)$.

REFERENCES

Bates, W. J., 1946, *Observatory*, **66,** 362; 1947, *Proc. Phys. Soc.*, **59,** 940.
Brown, D. S., 1954, *Proc. Phys. Soc.* B, **67,** 232.
Candler, C., 1951, *Modern Interferometers* (London : Hilger and Watts), Ch. VII.
Drew, R. L., 1951, *Proc. Phys. Soc.* B, **64,** 1005.
Gates, J. W., 1954, *Brit. J. Appl. Phys.*, **5,** 133.
Hopkins, H. H., 1950, *Wave Theory of Aberrations* (Oxford : Clarendon Press), p. 48.
Kingslake, R., 1926, *Trans. Opt. Soc., Lond.*, **27,** 94; **28,** 1.
Kösters, W., 1931, German Patent 4th March 1931, *Reichspatentschrift* 595211 (1934).

On the Basic Approximation in the Theory of the Phase Contrast Microscope

By F. D. KAHN

Department of Astronomy, Physical Laboratories, University of Manchester

Communicated by E. Wolf; MS. received 17th February 1955 *and in final form 7th July* 1955

Abstract. Maxwell's equations are used to find a necessary condition which must be satisfied in order that the phase contrast microscope may function satisfactorily. The method in the paper requires no restriction to be placed on the general shape of the object to be examined: in particular it need not have a grating-like structure.

---

§ 1. Introduction

The theory of the phase contrast microscope has been described in the past by a number of authors (e.g. Zernike 1942, Bennett *et al.* 1946, Françon 1950, Hopkins 1953). It works because, under suitable conditions, an object of variable refractive index (here called a phase object) produces approximately the same diffraction effects as a similar body of variable conductivity (an amplitude object). But under these conditions the light diffracted by the phase object is 90° in phase ahead of that diffracted by the amplitude object. If the undiffracted light which has passed through a phase object is artificially delayed by 270° in phase then the phase object will produce interference effects as though it were an amplitude object. When such a phase object is examined under a microscope the regions of large optical thickness will appear dark and those of small optical thickness bright.

In this paper Maxwell's equations are used to show that the effect occurs only when certain parameters are sufficiently small. Criteria will be found to determine whether only a small error has been made in the approximation.

§ 2. Maxwell's Equations

Let μ be the permeability, ϵ the dielectric constant and σ the conductivity of the given medium. Then Maxwell's equations are, in gaussian units,

$$\nabla \times \mathbf{H} = \frac{4\pi}{c}\mathbf{j} + \frac{1}{c}\frac{\partial}{\partial t}(\epsilon \mathbf{E})$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial}{\partial t}(\mu \mathbf{H}) \qquad \ldots\ldots(1)$$

where $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic intensities and $\mathbf{j}$ the current density. By Ohm's law

$$\mathbf{j} = \sigma \mathbf{E}. \qquad \ldots\ldots(2)$$

For monochromatic radiation of frequency $\omega/2\pi$, $\mathbf{E} = \mathfrak{E} e^{i\omega t}$ and $\mathbf{H} = \mathfrak{H} e^{i\omega t}$ so that

$$\nabla \times \mathfrak{H} = \frac{4\pi\sigma}{c}\mathfrak{E} + \frac{i\omega\epsilon}{c}\mathfrak{E} \qquad \ldots\ldots(3)$$

$$\nabla \times \mathfrak{E} = -\frac{i\omega\mu}{c}\mathfrak{H} \qquad \ldots\ldots(4)$$

and

$$\nabla \times \left(-\frac{c}{i\omega\mu}\nabla \times \mathfrak{E}\right) = \left(\frac{4\pi\sigma}{c} + \frac{i\omega\epsilon}{c}\right)\mathfrak{E}$$

or

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{E}\right) + \left(\frac{4\pi i\omega\sigma}{c^2} - \frac{\omega^2\epsilon}{c^2}\right)\mathfrak{E} = 0. \qquad \ldots\ldots(5)$$

One expects no absorption effects and no diffraction of light when $\sigma=0$ and when μ and ϵ are slowly varying functions of position. The coefficients in the equation for $\mathfrak{E}$ satisfy these conditions when there is no phase or amplitude object in the field.

The equation for $\mathfrak{E}$ then becomes

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{E}\right) - \frac{\omega^2\epsilon}{c^2}\mathfrak{E} = 0 \qquad \ldots\ldots(6)$$

with suitable boundary conditions to allow for the admission of light to the apparatus. Let the solution of (6) be $\mathfrak{E} = \mathfrak{E}_i(\mathbf{r})$, say.

When an amplitude object is introduced, the conductivity is no longer zero everywhere, but equals $\delta\sigma(\mathbf{r})$ in a certain region S of space. The equation for $\mathfrak{E}$ still has the solution $\mathfrak{E} = \mathfrak{E}_i(\mathbf{r})$ in the zeroth approximation, but the accurate solution is $\mathfrak{E} = \mathfrak{E}_i(\mathbf{r}) + \mathfrak{A}(\mathbf{r})$ where

$$\nabla \times \left[\frac{1}{\mu}\nabla \times (\mathfrak{E}_i + \mathfrak{A})\right] - \frac{\omega^2\epsilon}{c^2}(\mathfrak{E}_i + \mathfrak{A}) + \frac{4\pi i\omega\delta\sigma}{c^2}(\mathfrak{E}_i + \mathfrak{A}) = 0 \qquad \ldots\ldots(7)$$

or, by use of the fact that $\mathfrak{E} = \mathfrak{E}_i$ satisfies (6),

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{A}\right) - \frac{\omega^2\epsilon}{c^2}\mathfrak{A} = -\frac{4\pi i\omega\delta\sigma}{c^2}(\mathfrak{E}_i + \mathfrak{A}). \qquad \ldots\ldots(8)$$

The function $\mathfrak{A}(\mathbf{r})$ must also satisfy the Sommerfeld radiation condition (cf. Baker and Copson 1950).

Now let a phase object be introduced into the field instead of the amplitude object, then $\delta\sigma \equiv 0$ throughout the medium, and the dielectric constant becomes $\epsilon + \delta\epsilon$, where $\delta\epsilon = \delta\epsilon(\mathbf{r}) \neq 0$ in the region S. If $\delta\epsilon(\mathbf{r})$ here is everywhere proportional to $\delta\sigma(\mathbf{r})$ in the amplitude object then the two objects may be said to be similar. With the same boundary conditions as before $\mathfrak{E} = \mathfrak{E}_i(\mathbf{r})$ is still the solution of the equations in the zeroth approximation. The complete solution is now $\mathfrak{E} = \mathfrak{E}_i(\mathbf{r}) + \mathfrak{P}(\mathbf{r})$.

By an argument similar to that given for the amplitude object, we find for the phase object that

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{P}\right) - \frac{\omega^2\epsilon}{c^2}\mathfrak{P} = \frac{\omega^2\delta\epsilon}{c^2}(\mathfrak{E}_i + \mathfrak{P}). \qquad \ldots\ldots(9)$$

A simple approximate relation exists between $\mathfrak{A}$ and $\mathfrak{P}$ if, and only if, $|\mathfrak{A}|$ and $|\mathfrak{P}|$ are small compared with $|\mathfrak{E}_i|$, for then the terms $\mathfrak{E}_i + \mathfrak{A}$ and $\mathfrak{E}_i + \mathfrak{P}$ on the right-hand sides of (8) and (9) can both be replaced by $\mathfrak{E}_i$, and we have

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{A}\right) - \frac{\omega^2\epsilon}{c^2}\mathfrak{A} = -\frac{4\pi i\omega\delta\sigma}{c^2}\mathfrak{E}_i \qquad \ldots\ldots(10)$$

and

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathfrak{P}\right) - \frac{\omega^2\epsilon}{c^2}\mathfrak{P} = \frac{\omega^2\delta\epsilon}{c^2}\mathfrak{E}_i. \qquad \ldots\ldots(11)$$

Further, $\mathfrak{A}$ and $\mathfrak{P}$ satisfy the same boundary conditions. If $\delta\sigma$ and $\delta\epsilon$ are related by

$$\omega\delta\epsilon=4\pi\delta\sigma, \quad \text{i.e.} \quad \frac{2\pi c\delta\epsilon}{\lambda_0}=4\pi\delta\sigma \quad \text{or} \quad \frac{\delta\sigma}{\delta\epsilon}=\frac{c}{2\lambda_0} \qquad \ldots\ldots(12)$$

the amplitude and the phase object are equivalent. $\lambda_0(=2\pi c/\omega)$ is the wavelength *in vacuo* of the incident light.

Relation (12) shows that the objects can be made equivalent only for light of a given wavelength. Suppose two objects are equivalent at a wavelength λ_e then, at a wavelength $\lambda_0>\lambda_e$, the phase object will produce the same diffraction effects as an amplitude object whose conductivity is lower, in the ratio $\lambda_e:\lambda_0$, than that of the amplitude object which was equivalent to it at wavelength λ_e.

When the objects are equivalent the coefficients in equations (10) and (11) are the same, except for the factor $-i$ on the right-hand side of (10). It follows that $\mathfrak{A}=-i\mathfrak{P}$. But the electric vector of the light diffracted by the amplitude object is $\mathbf{A}=\mathfrak{A}e^{i\omega t}$ while the vector of the light diffracted by the phase object is $\mathbf{P}=\mathfrak{P}e^{i\omega t}=i\mathfrak{A}e^{i\omega t}=\mathbf{A}e^{i\pi/2}$. Therefore, to this approximation, the two objects produce identical diffracted waves, but in the case of the phase object the waves are 90° ahead of those diffracted by the amplitude object.

This approach is valid only when $|\mathfrak{E}_i|$ is much greater than $|\mathfrak{A}|$ and $|\mathfrak{P}|$ at points where $\delta\sigma$ or $\delta\epsilon$ is significantly different from zero. If the vectors $\mathfrak{A}$ and $\mathfrak{P}$ cannot be neglected on the right-hand sides of equations (10) and (11), then the simple relation between them does not hold and the basic approximation fails.

§ 3. THE VALIDITY OF THE APPROXIMATION

In the neighbourhood of the phase object both μ and ϵ will have, approximately, the constant values μ_0 and ϵ_0, and diffraction effects occur because within the region S the dielectric constant is $\epsilon_0+\delta\epsilon$. We shall consider first what happens when the incident light is a monochromatic, plane polarized plane wave. The effect of more general electromagnetic wave systems can be discussed afterwards by a superposition of the effects due to a series of such plane waves.

Then $\mathfrak{E}_i=\mathfrak{E}_0\exp(-i\mathbf{k}_0.\mathbf{r})$, and equation (11) becomes

$$-\nabla\times(\nabla\times\mathfrak{P})+\frac{\omega^2}{c^2}\mu_0\epsilon_0\mathfrak{P}$$

$$=\nabla^2\mathfrak{P}+\frac{\omega^2N^2}{c^2}\mathfrak{P}=-\frac{\omega^2}{c^2}\mu_0\delta\epsilon\mathfrak{E}_0\exp(-i\mathbf{k}_0.\mathbf{r}) \qquad \ldots\ldots(13)$$

where $N=(\mu_0\epsilon_0)^{1/2}$ is the refractive index of the background medium, if there are no space charges, and fluctuations in ϵ are large in linear extent compared with λ_0, so that $0=\nabla.(\epsilon_0+\delta\epsilon)\mathfrak{P}\simeq(\epsilon_0+\delta\epsilon)\nabla.\mathfrak{P}$. The incident light propagates through the medium with a speed $c/N=\omega/k_0$, and so $\omega N/c=k_0$. The wavelength of such light *in vacuo* is $\lambda_0=2\pi c/\omega=2\pi N/k_0$.

In the presence of the phase object the refractive index at a representative point $\mathbf{r}$ is changed to $(1+f)N$, where f is a function of $\mathbf{r}$, whose value is assumed to be small compared with unity. Then $N^2(1+f)^2=\mu_0(\epsilon_0+\delta\epsilon)$ or $2N^2f=\mu_0\delta\epsilon$. Equation (13) can now be written

$$\nabla^2\mathfrak{P}+k_0{}^2\mathfrak{P}=-\frac{2\omega^2N^2}{c^2}f(\mathbf{r})\mathfrak{E}_0\exp(-i\mathbf{k}_0.\mathbf{r})=-2k_0{}^2f(\mathbf{r})\mathfrak{E}_0\exp(-i\mathbf{k}_0.\mathbf{r}). \qquad \ldots\ldots(14)$$

We define $\phi(\mathbf{k})$ by the relation

$$f(\mathbf{r}) = \iiint_{-\infty}^{+\infty} \phi(\mathbf{k}) \exp(i\mathbf{k}.\mathbf{r})\, d\mathbf{k}. \qquad \ldots\ldots(15)$$

Then equation (14) can be written

$$\nabla^2 \mathfrak{P} + k_0{}^2 \mathfrak{P} = -2k_0{}^2 \mathfrak{E}_0 \iiint \phi(\mathbf{k}) \exp\{i(\mathbf{k}-\mathbf{k}_0).\mathbf{r}\}\, d\mathbf{k}$$

with the solution

$$\mathfrak{P} = -2k_0{}^2\mathfrak{E}_0 \iiint \frac{\phi(\mathbf{k}) \exp\{i(\mathbf{k}-\mathbf{k}_0).\mathbf{r}\}}{k_0{}^2 - (\mathbf{k}-\mathbf{k}_0).(\mathbf{k}-\mathbf{k}_0)}\, d\mathbf{k}$$

$$= 2k_0{}^2\mathfrak{E}_0 \exp(-i\mathbf{k}_0.\mathbf{r}) \iiint \frac{\phi(\mathbf{k}) \exp(i\mathbf{k}.\mathbf{r})}{k^2 - 2\mathbf{k}.\mathbf{k}_0}\, d\mathbf{k}. \qquad \ldots\ldots(16)$$

It is understood that the principal value of the improper integral is to be taken in (16). Before using (16) to make the comparison between $|\mathfrak{P}|$ and $|\mathfrak{E}_0|$ we must work out some properties of the function $\phi(\mathbf{k})$. Firstly, the origin O, where $\mathbf{r}=0$, can be chosen arbitrarily.[†] Secondly, since $f(\mathbf{r})$ is real everywhere, $f(\mathbf{r}) = f^*(\mathbf{r})$, and this holds for all $\mathbf{r}$.

Then

$$\iiint \phi(\mathbf{k}) \exp(i\mathbf{k}.\mathbf{r})\, d\mathbf{k} = \iiint \phi^*(\mathbf{k}) \exp(-i\mathbf{k}.\mathbf{r})\, d\mathbf{k}$$

$$= \iiint \phi^*(-\mathbf{k}) \exp(i\mathbf{k}.\mathbf{r})\, d\mathbf{k}$$

for all $\mathbf{r}$, so that $\phi(\mathbf{k}) = \phi^*(-\mathbf{k})$, for all $\mathbf{k}$.

We need to find the integral of $f(\mathbf{r})$ along a path parallel to the incident light. Let Oz be parallel to $\mathbf{k}_0$. Then we can write (15) in terms of polar coordinates k, θ, ψ in wave-number space, or, better, in terms of k, μ, ψ, where $\mu \equiv \cos\theta$.[‡] We find

$$f(z,0,0) \equiv f(z) = \int_0^\infty \int_{-1}^1 \int_0^{2\pi} \phi(k,\mu,\psi) \exp(ik\mu z)\, k^2 dk d\mu d\psi. \qquad \ldots\ldots(17)$$

Since $\phi(\mathbf{k}) = \phi^*(-\mathbf{k})$, it follows that $\phi(k,\mu,\psi) = \phi^*(k, -\mu, \pi+\psi)$, and after some manipulation

$$f(z) = \int_0^\infty \int_{-1}^1 \int_0^{2\pi} \phi^*(k,\mu,\psi) \exp(-ik\mu z)\, k^2 dk d\mu d\psi. \qquad \ldots\ldots(18)$$

Addition of (17) and (18) gives that

$$f(z) = \tfrac{1}{2} \int_0^\infty \int_{-1}^1 \int_0^{2\pi} [\phi(k,\mu,\psi) + \phi^*(k,\mu,\psi)] \cos(k\mu z) k^2 dk d\mu d\psi$$

$$+ \tfrac{1}{2} i \int_0^\infty \int_{-1}^1 \int_0^{2\pi} [\phi(k,\mu,\psi) - \phi^*(k,\mu,\psi)] \sin(k\mu z) k^2 dk d\mu d\psi$$

$$= \int_0^\infty \int_{-1}^1 \int_0^{2\pi} [\rho(k,\mu,\psi) \cos k\mu z - \tau(k,\mu,\psi) \sin k\mu z] k^2 dk d\mu d\psi \qquad \ldots\ldots(19)$$

where $\phi = \rho + i\tau$, and ρ and τ are real. Note that $\rho(k,\mu,\psi) = \rho(k, -\mu, \pi+\psi)$ while $\tau(k,\mu,\psi) = -\tau(k, -\mu, \pi+\psi)$.

Let

$$2\pi R(k,\mu) = \int_0^{2\pi} \rho(k,\mu,\psi)\, d\psi \quad \text{and} \quad 2\pi T(k,\mu) = \int_0^{2\pi} \tau(k,\mu,\psi)\, d\psi.$$

[†] But see Appendix.

[‡] μ is now used for $\cos\theta$. No confusion should arise with its use above for permeability.

Then $R(k,\mu)$ is an even function and $T(k,\mu)$ an odd function of μ. Further

$$f(z)=2\pi\int_0^\infty\int_{-1}^1[R(k,\mu)\cos k\mu z-T(k,\mu)\sin k\mu z]k^2dkd\mu$$

and so

$$\int_{-\infty}^z f(z)\,dz=2\pi\left[\int_0^\infty\int_{-1}^1\left\{\frac{R(k,\mu)\sin k\mu z}{\mu}+\frac{T(k,\mu)\cos k\mu z}{\mu}\right\}kdkd\mu\right]_{-\infty}^z. \quad \ldots\ldots(20)$$

The factor $1/\mu$ in the integrands introduces no difficulty near $\mu=0$, as long as $R(k,\mu)$ is finite and $T(k,\mu)$ is differentiable there, for all k. For we have $(\sin k\mu z)/\mu\to kz$; and $[T(k,\mu)]/\mu\to[\partial T(k,\mu)/\partial\mu]_{\mu=0}$ as $\mu\to 0$, since $T(k,\mu)$ is an odd function of μ and so $T(k,0)=0$.

Further

$$\lim_{z\to\infty}\int_{-1}^1\frac{R(k,\mu)\sin k\mu z}{\mu}d\mu=\lim_{z\to\infty}\int_{-kz}^{kz}R\left(k,\frac{\nu}{kz}\right)\frac{\sin\nu}{\nu}d\nu=\pi R(k,0)$$

and

$$\lim_{z\to-\infty}\int_{-1}^1\frac{R(k,\mu)\sin k\mu z}{\mu}d\mu=-\pi R(k,0). \quad \ldots\ldots(21)$$

Also

$$\lim_{z\to+\infty}\int_{-1}^1\frac{T(k,\mu)\cos k\mu z}{\mu}d\mu=0$$

provided only that $[T(k,\mu)]/\mu$, which is an even function of μ, finite for $\mu=0$, can be expressed as a convergent Fourier series in multiples of $\pi\mu$ over the range $-1<\mu<1$.

Thus, finally,

$$\int_{-\infty}^0 f(z)dz=2\pi\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}kdkd\mu+2\pi^2\int_0^\infty R(k,0)\,kdk$$

and

$$\int_{-\infty}^\infty f(z)\,dz=4\pi^2\int_0^\infty R(k,0)\,kdk. \quad \ldots\ldots(22)$$

The additional optical path along which the light travels from $-\infty$ to the point z in the presence of the phase object is

$$l(z)=N\int_{-\infty}^z f(z)\,dz.$$

The total additional optical path through the whole object is

$$4\pi^2N\int_0^\infty R(k,0)kdk=2L,\text{ say,}$$

so

$$l(0)=N\int_{-\infty}^0 f(z)dz=2\pi N\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}kdkd\mu+L$$

or

$$\frac{l(0)-L}{N}=2\pi\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}kdkd\mu. \quad \ldots\ldots(23)$$

Here $l(0)$ is the additional optical path to the point O, where $z=0$. The integral will occur again shortly.

Now the solution of the equations for $\mathfrak{P}$ is given in (16). We shall compare $\mathfrak{P}$ and $\mathfrak{E}_0$ at the point O ($\mathbf{r}=0$), remembering that this point can be chosen arbitrarily. In terms of k, μ and ψ

$$\mathfrak{P} = 2k_0^2\mathfrak{E}_0 \exp(-i\mathbf{k}_0\cdot\mathbf{r})\int_0^\infty\int_{-1}^1\int_0^{2\pi}\frac{\phi(k,\mu,\psi)}{k^2-2kk_0\mu}k^2dkd\mu d\psi$$

$$=4\pi k_0^2\mathfrak{E}_0\exp(-i\mathbf{k}_0\cdot\mathbf{r})\int_0^\infty\int_{-1}^1\frac{R(k,\mu)+iT(k,\mu)}{k-2k_0\mu}kdkd\mu. \qquad \ldots\ldots(24)$$

Since $R(k,\mu)$ is even and $T(k,\mu)$ is odd in μ

$$J=\int_{-1}^1\frac{R(k,\mu)+iT(k,\mu)}{k-2k_0\mu}d\mu=\int_{-1}^1\frac{R(k,\mu)-iT(k,\mu)}{k+2k_0\mu}d\mu$$

$$=k\int_{-1}^1\frac{R(k,\mu)}{k^2-4k_0^2\mu^2}d\mu+2ik_0\int_{-1}^1\frac{\mu T(k,\mu)}{k^2-4k_0^2\mu^2}d\mu. \qquad \ldots\ldots(25)$$

We must now restrict ourselves to phase objects whose typical linear dimensions are large compared with the wavelength of the incident light. Then, by the general properties of Fourier transforms, $R(k,\mu)$ and $T(k,\mu)$ are certainly significant only for values of $k \ll k_0 (\equiv 2\pi N/\lambda_0)$. The first term in (25) can then be neglected, with an error of the order of $\bar{k}/k_0$, where $\bar{l}=2\pi/\bar{k}$ is a typical length in the phase object.

Thus

$$J\simeq 2ik_0\int_{-1}^1\frac{\mu T(k,\mu)}{k^2-4k_0^2\mu^2}d\mu$$

$$=-\frac{i}{2}\int_{-1}^1\left[\frac{T(k,\mu)}{k+2k_0\mu}-\frac{T(k,\mu)}{k-2k_0\mu}\right]d\mu=-\frac{i}{2k_0}\int_{-1}^1\frac{T(k,\mu)}{\mu-k/2k_0}d\mu$$

since $T(k,\mu)=-T(k,-\mu)$. It follows that

$$I=\int_0^\infty J\,kdk=-\frac{i}{2k_0}\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu-k/2k_0}\,kdkd\mu\simeq-\frac{i}{2k_0}\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}kdkd\mu$$

provided that the distance of O from the centre of the phase object is much less than $l_c=N\bar{l}^2/\lambda_0$. This statement is justified in the Appendix.

In fact

$$I=-\frac{i}{2k_0}\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}k\,dk\,d\mu$$

with another error of order $\bar{k}/k_0$, and finally

$$\mathfrak{P}=-\frac{4\pi ik_0^2}{2k_0}\mathfrak{E}_0\exp(-i\mathbf{k}_0\cdot\mathbf{r})\int_0^\infty\int_{-1}^1\frac{T(k,\mu)}{\mu}k\,dk\,d\mu$$

or, by (23),

$$\mathfrak{P}=-ik_0\mathfrak{E}_0\exp(-i\mathbf{k}_0\cdot\mathbf{r})\frac{l(0)-L}{N}. \qquad \ldots\ldots(26)$$

But $2\pi N/k_0=\lambda_0$ and $\mathfrak{E}_i=\mathfrak{E}_0\exp(-i\mathbf{k}_0\cdot\mathbf{r})=$ undisturbed wave, so that

$$\mathfrak{P}=-i\mathfrak{E}_i 2\pi[l(0)-L]/\lambda_0. \qquad \ldots\ldots(27)^\dagger$$

Thus $\mathfrak{P}=0$ at the point on the incident ray where $l(0)=L$, which is so placed that the additional optical path from $-\infty$ to the point equals the additional

---

[†] The properties of the magnetic vector can be deduced from (27) by means of the relation $\nabla\times\mathfrak{E}=-i\omega\mu\mathfrak{H}/c$ (μ is the permeability).

optical path from the point to $+\infty$. We call this point the optical centre of the object for the given ray.

In general equation (27) shows that, to the first approximation, *the diffracted electric vector* $\mathfrak{P}$ *is parallel to the undisturbed vector* $\mathfrak{E}_i$, *but is* 90° *out of phase with it*, at points distant much less than l_c from the optical centre. Here $|\mathfrak{P}|$ and $|\mathfrak{E}_i|$ are in the same ratio as the length of the additional optical path from the optical centre to the point considered and $1/2\pi$ times the vacuum wavelength of the incident light. The basic approximation holds if $|\{l(0)-L\}/\lambda_0| \ll 1$ at all points in the field, i.e. if the phase object introduces an additional optical path which is small, compared with the vacuum wavelength of the light used, along any ray parallel to the incident light.

The result can now be generalized by a superposition of a number of incident plane waves, and we find the following:

A phase object of mean additional refractive index n will, when examined under a phase microscope with light in a spectral range $\Delta\lambda$ around the wavelength λ_0, produce approximately the same diffraction effects, advanced by 90° in phase, as an amplitude object equivalent to it for light of wavelength λ_0. The error made through equating the diffraction effects in the two cases is of the order of the largest of the three ratios $\Delta\lambda/\lambda_0$, $\lambda_0/\bar{l}$ and $n\mathscr{L}/\lambda_0$, where $\bar{l}$ is the linear scale of the smallest fluctuations in the structure of the object and $\mathscr{L}$ is its maximum thickness along any incident ray.

ACKNOWLEDGMENTS

The author thanks Dr. E. Wolf most sincerely for much valuable discussion and for guiding him through the literature of the subject.

The author was a Turner–Newall Research Fellow at Manchester University at the time when this work was done.

APPENDIX

We shall discuss the condition under which one may write

$$\int_0^\infty \int_{-1}^{1} \frac{T(k,\mu)}{\mu-\nu}\, k\, dk\, d\mu \simeq \int_0^\infty \int_{-1}^{1} \frac{T(k,\mu)}{\mu}\, k\, dk\, d\mu$$

for small values of ν. A rigorous derivation of the condition would be out of place here. It will be enough to indicate why the approximation is good only as long as the point O is not too far from the centroid of the distribution of refracting material.

By definition

$$f(\mathbf{r}) = \iiint_{-\infty}^{\infty} \phi(\mathbf{k}) \exp(i\mathbf{k}\,.\,\mathbf{r})\, d\mathbf{k}$$

so that, on inversion,

$$\phi(\mathbf{k}) = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} f(\mathbf{r}) \exp(-i\mathbf{k}\,.\,\mathbf{r})\, d\mathbf{r}.$$

With cylindrical coordinates z, ρ, χ through the origin O in real space, and polar coordinates k, θ, ψ in wave-number space, we have

$$\phi(k,\mu,\psi) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \int_0^\infty \int_0^{2\pi} f(z,\rho,\chi) \exp\{-ik[z\mu + \rho(1-\mu^2)^{1/2} \cos(\chi-\psi)]\}\, dz\, \rho d\rho\, d\chi$$

where $\mu = \cos\theta$.

On integration over ψ we find

$$\int_0^{2\pi} \phi(k,\mu,\psi)\,d\psi = 2\pi[R(k,\mu)+iT(k,\mu)]$$

$$= \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{2\pi} f(z,\rho,\chi)\exp(-ikz\mu)\,dz\,\rho\,d\rho\,d\chi$$

$$\times \int_0^{2\pi} \exp\{-ik\rho(1-\mu^2)^{1/2}\cos(\chi-\psi)\}\,d\psi.$$

Now let
$$\int_0^{2\pi} f(z,\rho,\chi)\,d\chi = 2\pi F(z,\rho);$$

then, since
$$2\pi J_0(x) = \int_0^{2\pi} \exp\{-ix\cos(\theta-\alpha)\}\,d\theta,$$

it follows that

$$R(k,\mu)+iT(k,\mu) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_0^{\infty} F(z,\rho)\exp(-ikz\mu)J_0[k\rho(1-\mu^2)^{1/2}]\,dz\,\rho\,d\rho.$$

On taking imaginary parts

$$T(k,\mu) = -\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_0^{\infty} F(z,\rho)J_0[k\rho(1-\mu^2)^{1/2}]\sin(kz\mu)\,dz\,\rho\,d\rho.$$

If z_0 is the z coordinate of the centroid of the distribution of diffracting material then one can write $T(k,\mu)=G(k,\mu)\sin k\mu z_0$, where $G(k,\mu)$ varies only slowly with μ. For large values of kz_0 the rapid variations are all contained in the $\sin k\mu z_0$ term. In such a case

$$\int_0^{\infty}\int_{-1}^{1}\frac{T(k,\mu)}{\mu-\nu}\,k\,dk\,d\mu = \int_0^{\infty}\int_{-1}^{1}\frac{T(k,\mu+\nu)}{\mu}\,k\,dk\,d\mu + \text{other smaller terms}$$

$$= \int_0^{\infty}\int_{-1}^{1}\frac{G(k,\mu)\sin kz_0(\mu+\nu)}{\mu}\,k\,dk\,d\mu$$

$$+ \text{other smaller terms}$$

$$= \int_0^{\infty}\int_{-0}^{1}\frac{G(k,\mu)\sin kz_0\mu}{\mu}\cos(kz_0\nu)\,k\,dk\,d\mu$$

$$+ \text{other smaller terms}$$

$$\simeq \int_0^{\infty}\cos(kz_0\nu)\,k\,dk\int_{-1}^{1}\frac{T(k,\mu)}{\mu}\,d\mu.$$

The approximation holds only if $1-\cos kz_0\nu$ is small, or $kz_0\nu \ll \pi$ throughout the range where the integrand is significant. In this application $\nu = k/2k_0$ and the condition is $k^2z_0/k_0 \ll 2\pi$. If $\bar{l}$ is the typical length of the fluctuations in the diffracting object, the mean value of k is $\bar{k}=2\pi/\bar{l}$. Finally, since the vacuum wavelength of the incident light is $\lambda_0 = 2\pi N/k_0$ the condition becomes $\lambda_0 z_0/N\bar{l}^2 \ll 1$, or alternatively, $z_0/\bar{l} \ll N\bar{l}/\lambda_0$. Now the approximation needs to hold only at points within the diffracting object. The method in this paper therefore applies only if the ratio of the width of the object in the direction of the incident light to the linear extent of the typical fluctuations in the object is much smaller than N times the ratio of the extent of these fluctuations to the wavelength of the light used, where N is the refractive index of the background medium.

REFERENCES

Baker, B. B., and Copson, E. T., 1950, *Huygens' Principle*, 2nd Ed. (Oxford : University Press), Ch. 1.

Bennett, A. H., Jupink, H., Osterberg, H., and Richards, O. W., 1946, *Trans. Amer. Microscop. Soc.*, **65**, 99.

Françon, M., 1950, *Le contraste de phase* (Paris : *La Revue d'Optique*).

Hopkins, H. H., 1953, *Proc. Phys. Soc.* B, **66**, 331.

Zernike, F., 1942, *Physica*, **9**, 686.

Refractive Index Determinations by the Central Illumination (Becke Line) Method

By R. C. FAUST
British Rayon Research Association, Manchester

MS. received 16*th December* 1954, *and in amended form* 25*th April* 1955

Abstract. Refractive index determinations by the central illumination (or Becke line) method depend upon the diffraction of light by the specimen. The accuracy of the method is therefore influenced by the specimen dimensions and by the degree of defocusing of the microscope. Moreover, the method does not necessarily measure the refractive index of the surface regions of an inhomogeneous specimen.

It is also shown that it is possible to observe Fresnel diffraction bands with a spacing less than the conventional resolution limit. Consequently, such band spacings do not provide a reliable estimate of the resolution limit attainable in either a light or an electron microscope.

§ 1. INTRODUCTION

THE refractive index of a small transparent solid is usually determined by finding an immersion liquid of the same refractive index, the index of this liquid then being measured on a suitable refractometer. The most common method of detecting when the indices of the solid and liquid are equal is that of central illumination microscopy. In this method the immersed solid is illuminated by a narrow axial cone of light with the result that, when the microscope is defocused, a fringe pattern is seen at the edges of the specimen; under the conditions normally employed this pattern degenerates into a single bright band, the so-called Becke line. The index of the liquid is then changed until the fringe pattern disappears, whereupon it is usually safe to conclude that the index of the liquid is equal to that of the solid. (Not every disappearance corresponds to equality of refractive indices (Faust 1951).)

Although Spangenberg (1921, 1922) suggested that the Becke line arose from the diffraction of light, the explanation given in modern textbooks is based upon geometrical optics. The only exception known to the writer is the book by Rinne and Berek (1934), in which it is correctly stated that the Becke line can be regarded as arising from an asymmetrically diffracted wave originating at the specimen edge. This asymmetry occurs whenever the edge EE′ of a homogeneous object is inclined to the incident light direction (figure 1 (*a*)), its sense usually being such that the major part of the diffracted light is directed towards the medium of higher refractive index. Consequently, if the plane pp on which the microscope is focused lies above the object (figure 1 (*a*)), the fringe system on the high index side of the edge will be of greater contrast and visibility than that on the low index side, whereas if pp is below the object, the situation will be reversed.

When, however, the edge EE′ is parallel to the incident light direction (figure 1 (*b*)), the amplitude of the diffracted wave is symmetrical about the incident direction. (The disturbance in the diffracted wave is A at the diffraction angle α and $-A$ at the angle $-\alpha$, the change of sign denoting a phase difference of π (cf. equation (6 *a*)). Although the resulting fringe patterns are generally asymmetrical, they are, nevertheless, sufficiently well defined on both sides of the edge to be readily distinguishable from the patterns arising at an inclined edge (Faust 1951).†

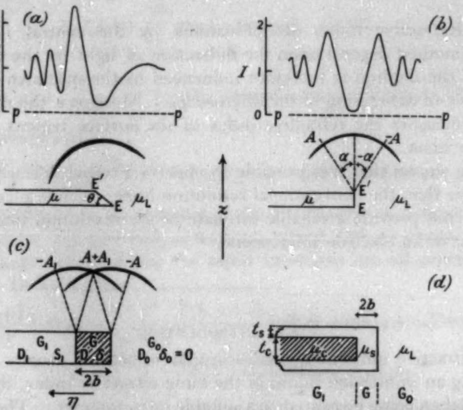


Figure 1. (*a*) The edge EE′ is inclined to the incident light direction (z axis), the asymmetrically diffracted wave being indicated by the variation of line thickness. On a plane pp above the object, the fringes are clearly visible on the high index side of the edge ($\mu > \mu_L$). (*b*) EE′ is parallel to the incident light direction and the fringes are clear on both sides of the edge. In both (*a*) and (*b*) the phase retardation of the sheet is π and the incident light intensity is unity. (*c*) Two symmetrically diffracted waves overlapping at a point above the centre of region G. (*d*) Object with a skin–core structure.

Two factors, the finite source size and the use of a sub-stage condenser, lead to the disappearance of the weaker diffraction fringes with the result that, when EE′ is inclined to the microscope axis, one is left with a single bright band, the Becke line, which lies on the high or low index side of the edge according to whether pp is above or below the object. If, on the other hand, EE′ is parallel to the axis, a relatively bright band is often found on both sides of the edge. The light in the outer regions of the incident cone does, however, fall upon EE′ at an angle, thereby inducing an asymmetry in the diffracted light, and this might then lead to the appearance of a single Becke line provided that the object is not too thin.

In this paper it is shown that the contrast of the diffraction pattern is influenced by the shape and dimensions of the specimen and by the degree of defocusing of the microscope, and that this in turn affects the accuracy with which the index of the immersion liquid can be matched to that of the solid. (Saylor (1935) had attempted to correlate the accuracy with the specimen dimensions, but his work must be regarded as incomplete because no reference was made to the degree of defocusing.) The paper also contains an examination of the common belief that the Becke line method measures the refractive index at the surface of an optically inhomogeneous specimen; examples of such specimens are biological and textile samples in which the molecular orientation changes from one region of the specimen to another.

---

† It has since been discovered that a similar study was undertaken by Obreimov (1945).

§ 2. THE FRESNEL–KIRCHHOFF TREATMENT

It is assumed that a thin object of arbitrary structure is situated in the plane $z=0$ and that it is illuminated by a parallel beam of monochromatic light propagated in the direction of the positive z axis (figure 2). An aberration-free objective LL is focused on the plane pp, which can be either above or below the object according to whether the defocusing z of the microscope is positive or negative. The plane p′p′, conjugate to plane pp, is the one that an observer will study by means of an eyepiece. The intensity distribution on p′p′ can be calculated from the Kirchhoff diffraction theory. A simplification is introduced if one assumes that the object structure is unidimensional: the disturbance at any point on the plane $z=0$ is then completely defined by the complex quantity $U_1(\eta)$, η being the coordinate of the point measured in a direction parallel to the plane of figure 2.

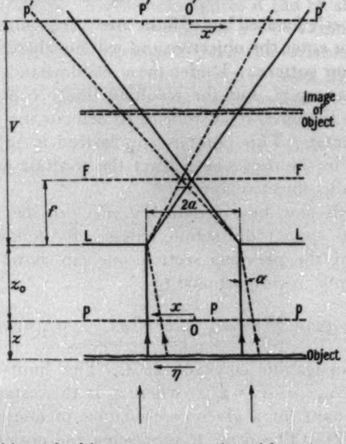


Figure 2. An object of arbitrary structure viewed by a lens of finite aperture.

All the light diffracted through the angle α will be brought to a focus on the rear focal plane FF, where a Fraunhofer diffraction pattern will be formed. From the calculated disturbance on FF one can find the disturbance $U'(x')$ at P′ (coordinate x') on the plane p′p′. When the lens aperture is infinite and the defocusing is zero, the expression for $U'(x')$ can be converted into the Fourier integral for $U_1(\eta)$ transformed on to the plane p′p′ (Menzel 1949). The image and the object are therefore identical in both amplitude and phase, and structural detail will remain undetected if the object influences only the phase of the incident light. If, however, the microscope is now defocused, any changes in the optical thickness in the specimen will be observed as intensity variations in the plane p′p′. Following Menzel (1951) one may then show that, for an infinite lens aperture, the disturbance at the point P′ is the same as that at the conjugate point P (coordinate x) on plane pp, that is,

$$_{a\to\infty}U'(x')=U(x)=\frac{e^{i\pi/4}}{\sqrt{2}}\int_{-\infty}^{\infty}U_1(\eta)\exp\left(-\tfrac{1}{2}i\pi v^2\right)dv \qquad \ldots\ldots(1)$$

where $v=(2/\lambda z)^{1/2}(\eta-x)$. Equation (1) is the usual expression for the Fresnel diffraction produced on the plane pp by an object at the distance z away.

The Kirchhoff theory is, however, open to a number of criticisms. It does not apply rigorously to diffraction at large angles nor, being scalar, does it explain any polarization effects. It is, moreover, assumed that the disturbance over an aperture in an opaque screen is the same as that occurring in the absence of the screen. Maréchal and Naudin (1950, 1951) have, however, found that near the aperture edge there is a region, only a few wavelengths wide, in which this assumption is not exactly fulfilled; analogous results have been found by workers in the field of microwaves. In view of these criticisms this theory must be applied with caution when the object is very narrow or when the defocusing z is very small.

§ 3. INFLUENCE OF LENS APERTURE

If the lens is properly focused but of finite aperture, some of the light diffracted by the object will not enter the objective and will not therefore contribute to the Fraunhofer diffraction pattern. Under these circumstances a phase object will be visible in the plane p′p′, but the resultant image contrast is usually small (Hopkins 1952). It is therefore customary to defocus the microscope in order to enhance the contrast. This point is emphasized in §§ 5 and 6, where it is shown that changes in the defocusing affect the contrast much more markedly than do changes in the objective aperture.

Consideration will now be given to the effect of the lens aperture on the resolution of closely spaced diffraction bands. By following the theoretical approach outlined in the previous section, one can show that the disturbance $U'(x')$ on the plane p′p′ is proportional to

$$\int_{-\infty}^{+\infty}U_2(\alpha)\,e^{-ikR}\,d\alpha=\int_{-\infty}^{+\infty}e^{-ikR}\,d\alpha\int_{-\eta_1}^{\eta_2}U_1(\eta)\,e^{ik\alpha\eta}\,d\eta \qquad \ldots\ldots(2)$$

where k is the propagation constant $2\pi/\lambda$. The limits of integration are $\eta_1=a+(z+z_0)\alpha$ and $\eta_2=a-(z+z_0)\alpha$, where z_0 is the distance between LL and pp, this being a constant for a given combination of microscope objective and eyepiece (cf. figure 2). The term R represents the optical path between the centre of the object ($\eta=0$) and a given field point x'; to a first approximation in α it is given by

$$R=f(x,z)+x\alpha \qquad \ldots\ldots(3a)$$

$$=f'(x',z)+\frac{x'}{V}a\cot\theta\sin\alpha \qquad \ldots\ldots(3b)$$

where θ is the semi-angular aperture of the objective, and V is the image distance LL–p′p′.

If the object is a semi-infinite sheet with an abrupt edge at $\eta=0$, the Fraunhofer diffraction pattern in the plane FF corresponds to that produced by a two-element echelon (Wood 1934, Darbyshire 1934), and the resultant diffraction angles are sufficiently small for one to write $\eta_1=\eta_2=a$. It can then be shown that, for a given phase retardation δ of the sheet, $U_2(\alpha)$ is determined by the value of the real number $m=a\sin\alpha/\lambda$. (For all values of δ the major part of the diffracted light is confined to the range $-1<m<1$.) Hence, if m, x' and z are fixed, an increase in the angular aperture θ of the objective will lead to a decrease in the second term of equation (3 *b*), and the disturbance $U'(x')$ on p′p′ will more closely approach the ideal disturbance $U(x)$ on plane pp.

Rayleigh (1899) has shown that an image can be regarded as sensibly perfect even if, as a result of aberrations, the phase at one extremity of the wave surface is a quarter-period in advance of, and that at the other extremity a quarter-period in the rear of, the phase at the centre. Consequently, the diffraction pattern observed on p′p′ should not sensibly differ from the ideal pattern on pp provided that, for a given field coordinate x', the second term of equation (3) is limited to the range $-\frac{1}{4}\lambda$ to $+\frac{1}{4}\lambda$.

This criterion offers an explanation of the experiment by Wegman (1950) in which he observed diffraction bands separated by less than half the theoretical resolution limit. The following information was given: the wavelength was 4916 Å, the defocusing was 0·25 mm, the smallest detectable band separation was 15 000 Å, and the numerical aperture of the objective was 0·13 yielding a resolution limit of 38 000 Å. From these figures it can be shown that, when the band separation was 15 000 Å, the coordinate $x=0{\cdot}007$ cm. As most of the diffracted light lay within the limits $\sin\alpha=\pm\lambda/a$, the variable term of equation (3) fell within the limits $\pm 0{\cdot}007\lambda/a$. The objective used by Wegman originally possessed a numerical aperture of 0·8 and was presumably of 4 mm focal length. As the numerical aperture was reduced to 0·13 by means of a diaphragm, the effective linear aperture a of the lens was approximately 0·5 mm. The second term of equation (3) was thus effectively restricted to the range $-\lambda/7$ to $+\lambda/7$, thereby satisfying the Rayleigh criterion. This argument provides theoretical justification for the statement by Wegman that the separation of the diffraction bands cannot be regarded as a measure of the practical resolution limit of an objective. (The writer has observed fringes that were so clearly resolved that there was no indication that their separation ($2{\cdot}3\mu$) was almost the same as the resolution limit ($2{\cdot}1\mu$) calculated from the numerical aperture of the objective.) The chief factors limiting the resolution of the diffraction bands are the finite spectral line width and the imperfect collimation of the incident radiation; in the electron microscope the lens aberrations and circuit stability are also important factors. (μ here indicates microns).

§ 4. THEORY APPLIED TO PHASE STRIPS

Two types of specimen are considered in this paper. The first is a phase strip with a rectangular cross section in the plane of figure 2; the other, which is dealt with in a later section, is that of a right cylinder with its axis normal to the plane of figure 2.

The theory of the phase strip is developed by considering an object consisting of three uniform regions $G_0(-\infty<\eta<-b)$, $G(-b<\eta<b)$ and $G_1(b<\eta<\infty)$. Each is of infinite extent in a direction perpendicular to the plane of the paper and each is separated from its neighbour by an edge parallel to the microscope axis and therefore parallel to the incident light direction (figure 1(*c*)). The disturbance on the plane $z=0$ immediately behind the object is thus given directly by $U_1(-\eta)=D_0$, $D\exp(-i\delta)$ and $D_1\exp(-i\delta_1)$, where D_0, D and D_1 are the amplitude reduction factors of the three regions and 0, δ and δ_1 are the phase retardations relative to the light passing through G_0.

From equation (1) the disturbance at a point $P(x,z)$ on the plane pp is

$$U(x)=D_0S^{\tau}_{-\infty}+D\exp(-i\delta)S^{\tau_1}_{\tau}+D_1\exp(-i\delta_1)S^{\infty}_{\tau_1} \qquad \ldots\ldots(4)$$

in which

$$S^{\beta}_{\alpha}=\frac{e^{i\pi/4}}{\sqrt{2}}\int_{\alpha}^{\beta}\exp(-\tfrac{1}{2}i\pi v^2)\,dv \qquad \ldots\ldots(5a)$$

and

$$\tau=-(2/\lambda z)^{1/2}(x+b), \quad \tau_1=-(2/\lambda z)^{1/2}(x-b). \qquad \ldots\ldots(5b)$$

For convenience equation (4) can be expressed in three equivalent forms. If the plane pp is above the object (i.e. if z is positive), these are

$$\begin{aligned} U(x)&=D_0-A+A_1 && x<-b\\ U(x)&=D\exp(-i\delta)+A+A_1 && -b<x<b\\ U(x)&=D_1\exp(-i\delta_1)+A-A_1 && x>b \qquad \ldots\ldots(6a)\end{aligned}$$

where

$$A=[D_0-D\exp(-i\delta)]S^{\infty}_{|\tau|}, \quad A_1=[D_1\exp(-i\delta_1)-D\exp(-i\delta)]S^{\infty}_{|\tau_1|}. \qquad \ldots\ldots(6b)$$

The terms A and A_1 represent symmetrical waves spreading from the edges G_0–G and G–G_1. The diffraction pattern can therefore be regarded as being formed by the superposition of these edge-waves upon the disturbance that would be present were light propagated according to the laws of geometrical optics.

By multiplying the disturbance $U(x)$ by its complex conjugate, one can calculate the intensity on the plane pp; this is also the intensity I_∞ observed on p′p′ when the microscope objective is of infinite aperture. Should the lens be focused on a plane pp below the object, z will be negative and the observed intensity can then be derived from the relationship

$$I(x,-|z|,\delta,\delta_1)=I(x,|z|,-\delta,-\delta_1). \qquad \ldots\ldots(7)$$

For a semi-infinite sheet ($D=D_1$ and $\delta=\delta_1$) the wave A_1 vanishes, leaving the wave A spreading from the point ($\eta=-b$, $z=0$). As the intensity at the point $P(x,z)$ is now determined by the magnitude of $(x+b)/(\lambda|z|)^{1/2}$, the diffraction pattern will be of the same form on all planes $z=$ constant, the scale of the pattern being proportional to $|z|^{1/2}$ (Obreimov 1945, Kinder and Recknagel 1947). When, however, two boundaries are present, the form of the pattern is not independent of the degree of defocusing but depends upon the magnitude of

$$\Delta\tau=\tau_1-\tau=(2/\lambda|z|)^{1/2}2b.$$

This is termed the reduced width because the form of the diffraction pattern will be identical for all objects with the same value of $\Delta\tau$, provided that the optical properties of the regions G_0, G and G_1 are unaltered.

In figure 3 are plotted the calculated intensity distributions for a rectangular strip G ($\delta=\pi$) immersed in a homogeneous medium ($\delta_0=\delta_1=0$). (The object is regarded as non-absorbing, i.e. $D_0=D=D_1=1$.) When $\Delta\tau$ is greater than about ten, the overlapping of the two edge-waves A and A_1 is of little significance because of the rapid decrease in the amplitudes of these waves with increasing angles of diffraction. The observed diffraction pattern will therefore consist of two band systems, each of which will have nearly the same appearance as that formed at

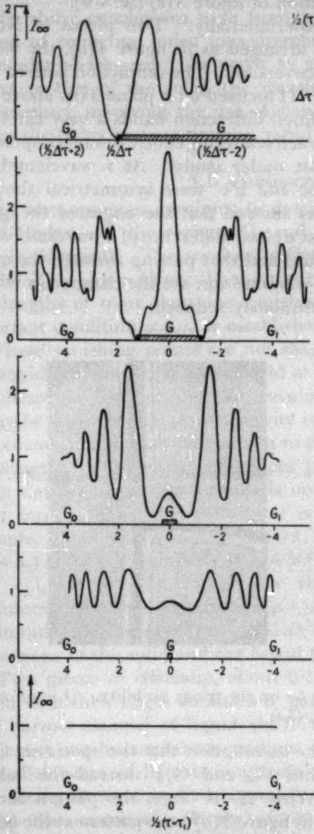


Figure 3. The calculated diffraction pattern of an isolated object of reduced width $\Delta\tau$ ($\delta_0=\delta_1=0$, $\delta=\pi$). The intensity I_∞ is plotted against

$$\tfrac{1}{2}(\tau+\tau_1)=-(2/\lambda|z|)^{1/2}x.$$

The incident intensity is taken as unity.

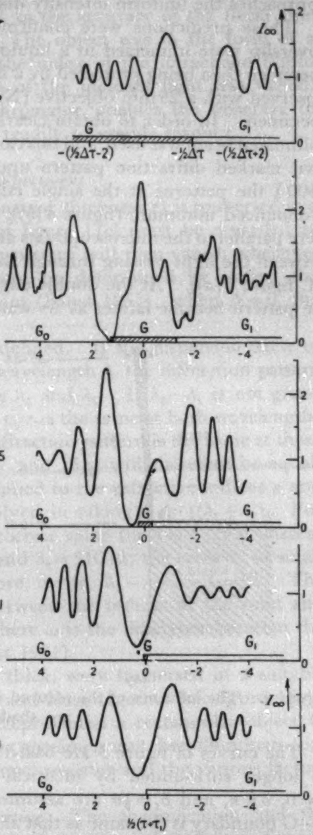
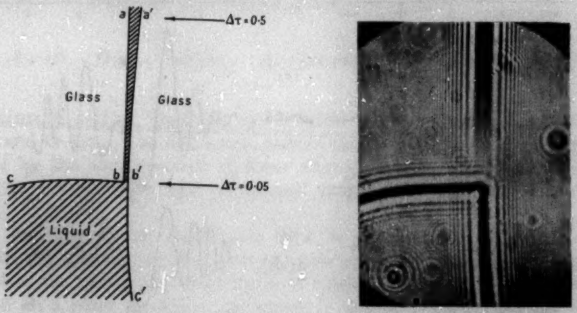


Figure 5. As in figure 3, but the object is now bordered by dissimilar media ($\delta_0=0$, $\delta=\pi$, $\delta_1=\tfrac{1}{2}\pi$).

the boundary between two semi-infinite sheets differing in their phase retardations by π (figure 3 (*a*)). As $\Delta\tau$ decreases, the overlapping of the edge waves becomes increaingly important and the pattern is strikingly modified (figures 3 (*b*) and (*c*)). When finally $\Delta\tau$ is very small, the diffraction pattern will be almost identical with that which would be produced at the boundary between two semi-infinite areas with the properties D_0, 0 and D_1, δ_1. This is illustrated in figure 3 (*d*) ($\Delta\tau = 0{\cdot}1$); since the regions G_0 and G_1 are identical, the pattern is of poor contrast and approaches the uniform intensity distribution of figure 3 (*e*) ($\Delta\tau = 0$).

These predictions were confirmed experimentally. Two pieces of No. 0 coverslip were immersed in a liquid and arranged as in figure 4 (*a*), the whole specimen then being protected by a large coverslip. The diffraction pattern was observed with a 25 mm objective (N.A. 0·21) focused on a plane 3 cm above the specimen. In order to obtain clearly defined diffraction bands a very carefully collimated beam was used, and this caused each scattering centre to superimpose its own marked diffraction pattern upon that under study. At a wavelength of 5400 Å the patterns at the single edges bc and b′c′ were symmetrical about a pronounced minimum (figure 4 (*b*)). This showed that the edges of the glass were parallel to the microscope axis and that a phase difference of π was established between the light passing through the liquid and that passing through the glass (cf. figure 3 (*a*)). At the double edge ab–a′b′ one can see the manner in which the pattern became fainter as $\Delta\tau$ was continuously reduced.



(*a*) (*b*)

Figure 4. The influence of the reduced width $\Delta\tau$ on the fringe visibility. (Liquid areas are shaded in (*a*.))

The curves of figure 5 are based on the assumption that the specimen G is no longer surrounded by identical regions G_0 and G_1; instead the values $\delta_0 = 0$, $\delta = \pi$, and $\delta_1 = \frac{1}{2}\pi$ are assumed. When $\Delta\tau$ is large, the pattern at the G_0–G boundary is the same as that shown in figure 3 (*a*); the pattern at the other boundary is identical with that which would occur were G a semi-infinite sheet introducing a phase retardation of $\frac{1}{2}\pi$ with respect to the region G_1 (figure 5 (*a*)). As $\Delta\tau$ is reduced the patterns at the two edges of the specimen merge until only one pronounced minimum can be observed (figure 5 (*d*)). Although the intensity distribution is not then the same as that obtaining when $\Delta\tau = 0$ (figure 5 (*e*)), the similarity is such that the two patterns could easily be confused when inspected visually.

Two coverslips were placed side by side in a liquid as in figure 4(*a*) but in this instance the slips, corresponding to regions G_0 and G_1, were of unequal thickness. It was then confirmed that, for small values of $\Delta\tau$, the diffraction pattern along the double edge ab–a′b′ was virtually determined by the relative properties of G_0 and G_1, the nature of region G exerting but little influence.

§ 5. INDEX MEASUREMENTS ON PHASE STRIPS

The above conclusions have been related to the accuracy of the Becke line method of refractive index measurements. The object was in the form of a rectangular phase strip. In order to alter the index of the immersion liquid relative to that of the solid, the wavelength of the incident light was varied. This method, involving the use of a monochromator, enables the observer to change the index in a continuous, rapid and readily reversible manner.

5.1. *Homogeneous Specimens*

If a homogeneous specimen (index μ, constant thickness t) is immersed in a liquid (index μ_L), the regions G_0 and G_1 of figure 1(*c*) must be regarded as optically identical. The resultant diffraction pattern will then vanish whenever $\delta=(2\pi/\lambda)(\mu-\mu_L)t=2N\pi$. The integer N can be determined by noting the wavelengths of more than one vanishing point (Faust 1951); when $N=0$, the important condition $\mu=\mu_L$ is established.

These vanishing points are not sharply defined. If the dispersion curve of the liquid intersects that of the solid at the wavelength λ, the diffraction pattern will disappear between the two wavelengths λ_1 and λ_2. If $\lambda_2-\lambda_1$ is not great, it may be assumed that the sensitivity of the eye is the same at both wavelengths and, consequently, that the contrast in the diffraction pattern is the same at these wavelengths. The phase differences $|\delta_{\lambda 1}|=$ and $|\delta_{\lambda 2}|$ will therefore be equal. If the Cauchy dispersion formula is now applied to the refractive indices μ and μ_L, it can be shown that little error is involved in taking λ as $\frac{1}{2}(\lambda_1+\lambda_2)$. For example, if $\lambda_1=5000$ Å and $\lambda_2=5500$ Å, the correct value for λ is 5225 Å whereas $\frac{1}{2}(\lambda_1+\lambda_2)$ is 5250 Å; similarly if $\lambda_1=5000$ Å and $\lambda_2=5100$ Å, the error is so small that $\frac{1}{2}(\lambda_1+\lambda_2)-\lambda=1$ Å. One can therefore write $\lambda_2-\lambda=\lambda-\lambda_1=d\lambda$. The minimum detectable difference $d(\mu-\mu_L)$ between the indices of the solid and the liquid is then proportional to $\omega\, d\lambda/\lambda^3$, where ω is the difference between the dispersions of the solid and the liquid (Faust 1951).

Two pieces of coverslip, about 0·1 mm thick, were immersed in a suitable liquid with their edges ab and a′b′ parallel and close together (cf. figure 4(*a*)). The narrow channel of liquid aba′b′ thus represented a rectangular object G immersed in a homogeneous medium. This specimen was then illuminated by a parallel beam of light issuing from a monochromator and the various factors influencing $d\lambda$ were studied separately.

(i) *The numerical aperture of the objective.* The magnification and the defocusing were kept constant and the width of the channel aba′b′ was fixed at 5 microns. It was then found that $d\lambda$ was independent of the numerical aperture in the tested range $0{\cdot}1<\text{N.A.}<0{\cdot}65$. This conclusion was also true for a single edge ($\Delta\tau=\infty$).

(ii) *The defocusing* z was then changed in order to alter the reduced width $\Delta\tau$, the channel width again being kept at 5 microns. The measurements were made with three different objectives, the total magnification being maintained at about 45 diameters. The results are plotted in figure 6, on the right-hand side of which is marked the minimum detectable refractive index difference $d(\mu-\mu_L)$ associated with a given wavelength range $d\lambda$. The wavelengths λ_1 and λ_2 were independent of the sign of z (cf. equation (7)), and the mean values of $d\lambda$ have therefore been plotted against $|z|$.

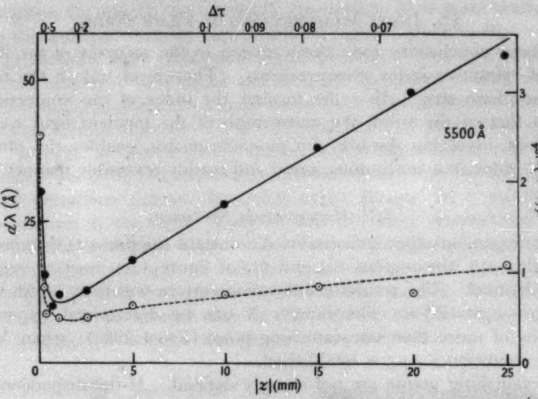


Figure 6. The wavelength uncertainty $d\lambda$ as a function of the defocusing $|z|$ for a specimen producing symmetrically diffracted waves.

The results for a single edge ($\Delta\tau=\infty$) lie on the dotted curve. On the side of decreasing $|z|$ the steep rise is caused by the fact that the pattern shrinks to such small lateral dimensions that it is barely visible. (Theoretically, the edge should be invisible when $z=0$, but the restricted lens aperture and the finite specimen thickness cause it to be seen as a faint, narrow line.) The gentle rise in the curve as $|z|$ increases is to be expected because the incident light is neither perfectly collimated nor strictly monochromatic; consequently, the increasing lateral spread of the pattern will result in a deterioration of contrast. The value of $|z|$ that leads to the minimum in the curve will therefore depend upon the magnification of the microscope, the wavelength composition of the incident light, and upon whether or not a condenser is employed.

The full curve applies to the rectangular specimen aba′b′. When $\Delta\tau$ is large (i.e. when $|z|$ is small) the waves from the two specimen edges overlap but slightly and $d\lambda$ will be the same as that for a single edge. As $\Delta\tau$ is reduced below 0·4 the two edge waves overlap to an increasing extent, the contrast declines rapidly and the range $d\lambda$ rises continuously (cf. figure 3).

(iii) *The object width 2b.* According to figure 3 the contrast of the diffraction pattern is approximately constant for values of $\Delta\tau$ above about 0·5; below this value the contrast gradually diminishes. However, because of the complications mentioned in the previous section, a plot of $d\lambda$ against $|z|$ produces a curve with a minimum. If, on the other hand, the defocusing is kept fixed whilst the object width is altered, then the resultant changes in $\Delta\tau$ should produce changes of contrast in accordance with figure 3. This is confirmed by the results quoted in the first three lines of the table; these were obtained with a fixed defocusing of $z=10$ mm, a mean wavelength of 5500 Å, and a lens with a numerical aperture of 0·15. To each value of $\Delta\tau$ it is possible to obtain from figure 6 a corresponding value of $d\lambda$ (fourth line). However, this value can be compared with that shown in the third line only when it falls in the range for which the results on the single edge (dotted curve) are but slightly dependent on the defocusing $|z|$. For this reason no comparison value is given in columns 2 and 3; the comparison value in the final column was obtained by extrapolation from figure 6.

$2b$ (microns)	∞	118	40	21	15	7	5	2·5
$\Delta\tau$	∞	2·24	0·75	0·40	0·28	0·13	0·09	0·05
$d\lambda$	10	7	8	7	10	24	35	72
$d\lambda$ (fig. 6)	11	—	—	10	10	18	28	87

(iv) *The specimen thickness.* In an earlier paper (Faust 1951) it was shown that $d\lambda$ was proportional to the thickness t. Although the full significance of the degree of focusing was not realized at that time, the results are valid because the degree of focusing was kept constant and the same objective and eyepiece were used throughout.

(v) *The wavelength.* The term $\Delta\tau$, depending as it does on $\lambda^{-1/2}$, will not be appreciably influenced by wavelength changes. The range $d\lambda$ for a rectangular specimen should therefore depend on the wavelength in the same manner as for a simple edge, namely that between 5000 Å and 6000 Å $d\lambda$ is roughly proportional to λ^4 (Faust 1951); this has been confirmed experimentally. The minimum detectable index difference is, however, proportional to $d\lambda/\lambda^3$ and will therefore alter but little throughout the visible spectrum.

5.2. *Inhomogeneous Specimens*

An example of such an object is provided by a semi-infinite sheet of index μ_C and thickness t_C surrounded by a skin of index μ_S, thickness t_S and width $2b$ (figure 1 (*d*)). This object is immersed in a liquid of index μ_L. The resulting diffraction pattern will always be symmetrical about the centre of the skin region G provided that

$$\delta_1=(2\pi/\lambda)[(\mu_C t_C+2\mu_S t_S)-\mu_L(t_C+2t_S)]=2n\pi, \qquad \ldots\ldots(8)$$

n being integral. If the position $n=0$ is found, the measured index μ_L will be intermediate between that of the skin and that of the core. Hence, knowing the index of the skin, one can discover whether μ_S or μ_C is the greater.

The successful determination of the skin index does, however, necessitate the fulfilment of certain conditions. If the reduced width $\Delta\tau=(2/\lambda|z|)^{1/2}2b$ of the skin is large, the patterns at the liquid–skin (G_0-G) and the skin–core $(G-G_1)$ boundaries will be readily distinguishable. Since the former pattern will disappear whenever $\delta=2N\pi$, the index of the skin can be determined by noting the disappearance corresponding to $N=0$ (i.e. $\mu_L=\mu_S$). If, however, $\Delta\tau$ is about 0·5, the patterns at the two boundaries will be barely distinguishable and a satisfactory determination of μ_S will not be practicable. If $\Delta\tau$ is still smaller (0·1 or less), the observed pattern will hardly be influenced by the properties of the skin region G. The pattern will therefore be of minimum contrast whenever $\delta_1=2n\pi$; as shown above this leads to a value of μ_L between μ_S and μ_C.

These considerations show that, if the index of the skin is to be measured, then the reduced width $\Delta\tau$ of the skin must not be much less than unity. This means that the defocusing must not exceed 16 microns for a geometrical skin width of 2 microns, or 4 microns for a skin 1 micron thick. Furthermore, not only must the defocusing be less than a certain value, but the numerical aperture of the objective must be adequate to resolve the patterns at the two boundaries of the skin. Consequently one cannot accept without qualification the widely held belief that the central-illumination method measures the surface properties of a specimen.

§ 6. INDEX MEASUREMENTS ON CYLINDERS

If, as was assumed in §§ 4 and 5, the changes in optical thickness are abrupt, the disturbance at any point on the plane $z=0$ is determined by the optical properties of the object at that particular point. No such simple approach is feasible with cylindrical and spherical objects. It is, however, possible to obtain a rigorous solution not from the Fresnel theory but from the electromagnetic theory of light. The result is usually expressed as an infinite series of Bessel functions but the convergence is extremely poor when the fibre radius is greater than the wavelength of the incident radiation (see Spohn 1920, Bouwkamp 1954).

Because of these theoretical difficulties a purely experimental approach has been adopted in order to discover to what extent the conclusions of the earlier sections will apply to objects with gradual changes in optical thickness. The test objects were glass fibres of different diameters (6, 13, 25 and 50 microns) but of the same refractive index and dispersion.

As a result of the lens action of the fibre it is not to be expected that the wavelengths λ_1 and λ_2, between which the fibre disappears, should be independent of the sign of the defocusing z. For example, a series of measurements on a 6μ diameter fibre led to the following results: $\lambda_1=5148\pm 2$ Å, $\lambda_2=5317\pm 2$ Å for $z=-0{\cdot}6$ mm, and $\lambda_1=5134\pm 3$ Å, $\lambda_2=5302\pm 3$ Å for $z=0{\cdot}6$ mm. Although the differences $\lambda_2-\lambda_1=2d\lambda$ agreed to within the experimental error, the mean wavelength was 14 Å shorter with the positive defocusing than with the negative defocusing. It is therefore advisable, when working to the highest accuracy, to make determinations with the microscope defocused in both the positive and negative senses.

(i) *The numerical aperture of the objective* played only a small role. At constant magnification ($\times 100$) and constant defocusing the range $d\lambda$ for a 6μ diameter fibre rose by less than 30% upon changing from a numerical aperture of 0·25 to one of 0·65. Between numerical apertures of 0·1 and 0·25 no change in $d\lambda$ was detected. Saylor (1935), using a 30μ diameter fibre, found that the index uncertainty $d(\mu-\mu_L)$ had almost doubled upon changing the aperture from 0·25 to 0·4, and had more than doubled when the aperture was increased to 0·85. Unfortunately these results must be regarded as inconclusive because there was no control of either the magnification or the degree of defocusing.†

(ii) *The defocusing* z. The presence in the field of view of an isolated object of arbitrary cross-sectional shape will produce a disturbance on a plane immediately behind the object. That is, there will be a region in this plane over which $U_1(\eta)\neq 1$ (cf. equation (1)). The area of this region will be finite although not necessarily the same as that of the object (for example, with cylinders the diffracted light can spread beyond the geometrical limits of the specimen). It can therefore be deduced from equation (1) that the contrast in the diffraction pattern will be very small when the defocusing is appreciable.

---

† This criticism can also be made of table 3 in the author's (1951) paper.

Three objectives with numerical apertures between 0·1 and 0·25 were used, and the magnification was held at ×55. The range $d\lambda$ was found to be independent of the sign of z and the mean values have therefore been plotted in figure 7. These curves display the same characteristics as those of figure 6.

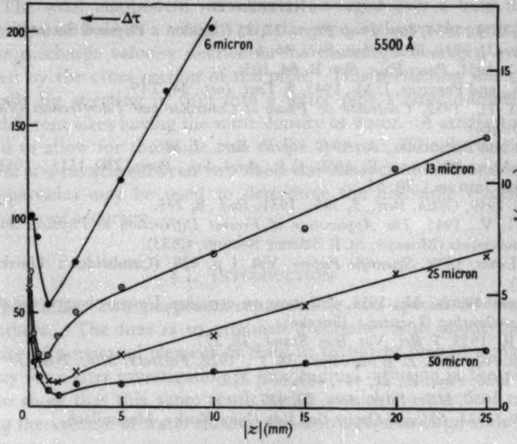


Figure 7. The wavelength uncertainty $d\lambda$ as a function of the defocusing $|z|$ for cylindrical specimens of different diameters.

(iii) *The fibre radius* ρ. The curves also show that the range $d\lambda$ increases as the fibre diameter decreases. Although there is no unique relationship between $d\lambda$ and ρ for all degrees of defocusing, $d\lambda$ is roughly proportional to $1/\rho$ for values of $|z|$ greater than 2 mm, that is, in the range for which the defocusing is relatively unimportant for a single edge (cf. figure 6).

(iv) *The wavelength.* It was again found that $d\lambda$ was approximately proportional to λ^4, from which it follows that the accuracy of the index measurements is nearly the same throughout the visible spectrum.

Inhomogeneous cylindrical specimens have been produced by coating glass fibres with thin collodion films. It was found that the measured refractive index depended upon the degree of defocusing and was not usually equal to that of the collodion film; with large defocusings the measured index did not even lie between the indices of the glass and the collodion. These findings are of importance in the textile field; they emphasize the conclusion of § 5.2 that the central illumination method does not necessarily measure the index at the surface of a specimen, and they provide a better understanding of the refractive index determinations made with the Preston fibre refractometer (Freeman and Preston 1943). It is hoped to publish details elsewhere.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Drs. H. H. Hopkins and G. A. J. Orchard for valuable help during the preparation of this paper. He also wishes to acknowledge his indebtedness to Messrs. Pilkington Brothers Ltd. for the supply of glass fibres of constant refractive index but different diameters.

This work forms part of the fundamental research programme undertaken by the British Rayon Research Association.

REFERENCES

BOUWKAMP, C. J., 1954, *Rep. Prog. Phys.*, **17,** 35 (London : Physical Society).
DARBYSHIRE, O., 1934, *Proc. Phys. Soc.*, **46,** 611.
FAUST, R. C., 1951, *Proc. Phys. Soc.* B, **64,** 105.
FREEMAN, K., and PRESTON, J. M., 1943, *J. Text. Inst.*, **34,** T19.
HOPKINS, H. H., 1952, *Contraste de Phase et Contraste par Interférences* (Paris : Revue d'Optique).
KINDER, E., and RECKNAGEL, A., 1947, *Optik, Berl.*, **2,** 346.
MARÉCHAL, A., and NAUDIN, F., 1950, *C.R. Acad. Sci., Paris*, **230,** 1515; 1951, *Rev. Opt.* (*Théor. Instrum.*), **30,** 65.
MENZEL, E., 1949, *Optik, Berl.*, **5,** 385; 1951, *Ibid.*, **8,** 295.
OBREIMOV, I. V., 1945, *The Application of Fresnel Diffraction to Physical and Technical Measurements* (Moscow, SCR Science Section, C633).
RAYLEIGH, LORD, 1879, *Scientific Papers*, Vol. 1 p. 429, (Cambridge : University Press, 1899).
RINNE, F., and BEREK, M., 1934, *Anleitung zu optischen Untersuchungen mit dem Polarisationsmikroskop* (Leipzig : Jänecke).
SAYLOR, C. R., 1935, *J. Res. Nat. Bur. Stand.*, **15,** 277.
SPANGENBERG, K., 1921, *Z. Wiss. Mikr.*, **38,** 1 ; 1922, *Fortschr. Min.*, **7–8,** 1.
SPOHN, H., 1920, *Phys. Z.*, **21,** 444, 469, 501, 518.
WEGMAN, L., 1950, *Helv. Phys. Acta*, **23,** 437.
WOOD, R. W., 1934, *Physical Optics*, 3rd Edn (New York : Macmillan).

The Mean Velocity of Discrete Particles in Turbulent Flow in a Pipe

By G. K. BATCHELOR, A. M. BINNIE and O. M. PHILLIPS
Trinity College, Cambridge

MS. received 13*th June* 1955

Abstract. The axial component of velocity, averaged over a long time, of a material element of fluid in a straight circular pipe is shown theoretically to be equal to the discharge velocity defined as the discharge, averaged over a long time, divided by the cross section of the pipe. This prediction was confirmed experimentally by injecting in turn into a water pipe a large number of solid spheres of different sizes having the same density as water. A satisfactory theory is developed to allow for the finite size of the spheres. The fluctuation in the time of travel of a particle between two fixed stations is examined, and it is shown how a few particles may be used to determine the discharge velocity within calculable limits of accuracy.

§ 1. Introduction

This paper has two purposes, representing the special interests of the co-authors. The first is to establish a certain result about the average velocity of a material element of fluid in turbulent flow in a pipe, and to pave the way for a later investigation of longitudinal diffusion in the pipe. The second is to show that this same result makes available a practical method of determining the volume of water flowing through a pipe in large-scale hydraulic installations.

The first of these two purposes is concerned essentially with the relation between two different ways of averaging the velocity of the fluid in the pipe. One way involves Eulerian quantities, namely, the averages over a long time of the fluid velocity at various fixed points, and the other Lagrangian quantities, which are associated with material elements of the fluid. In § 2 we shall give a plausible argument to show that the velocity of any material element of the fluid, averaged over a long time, is the same as what we shall call the discharge velocity, defined as the volume flux at some cross section, averaged over a long time, divided by the cross-sectional area. This equality of the two very different kinds of average velocity is the central result of the paper. In § 4 we describe experiments at the Engineering Laboratory, Cambridge, that confirm the result. The measurements of the Lagrangian average were not, of course, made with material fluid elements, but with solid spheres of the same density as the fluid and of various sizes, the extrapolation to zero size being made with the aid of a simple theory described in § 5.

The second of the two purposes arises from the fact that in a test of a large water turbine the accuracy of the hydraulic flow measurements is usually much lower than that of the electrical. One method of determining the flow that has been used successfully, and that is related in principle to the method suggested here, is to inject into the pipe salt solution or radioactive liquid—anything miscible, of which the concentration in water at points downstream may be measured—and to determine the time of arrival, at some position far downstream, of the maximum concentration of the injected liquid. The average velocity with which the position of maximum concentration has travelled is then taken as equal to the discharge velocity (Allen and Taylor 1923, Taylor 1954). Owing to the considerable dispersion of the injected liquid in the longitudinal direction it is necessary to observe continuously the concentration at some fixed point in order to determine the time of arrival of the maximum, and Allen's salt-velocity method is not simple in practice. We have thought it worth while to explore the possibility of an alternative method, which involves observations of the time taken by small solid particles to traverse a known length of pipe. In the last section of the paper, the method is assessed in the light of our experiments, and a numerical example shows that the accuracy of this method of determining the discharge velocity is good.

§ 2. THE RELATION BETWEEN THE MEAN VELOCITY OF A SINGLE PARTICLE AND THE DISCHARGE VELOCITY

If a steady pressure gradient is applied to a long length of pipe, the resulting turbulent motion in the fluid is statistically steady and independent of distance along the pipe. Material elements of fluid can wander in the transverse plane over a finite area only, and the velocity of a material element, $\mathbf{v}(t)$ say, is necessarily a stationary random function of time t as soon as the influence of the special choice of the element's position in the transverse plane at the moment of release has been lost. This stationary random function has a non-zero mean which we wish to determine.

Now it is an exact result of ergodic theory that, when a random function $\mathbf{v}(t)$ is stationary with respect to t, the probability average (sometimes called the ensemble or stochastic average) of $\mathbf{v}(t)$ is equal, under fairly general conditions, to the average obtained from an integral over a large range of values of t. Hence, the probability average of the velocity of a material element of fluid in the pipe (obtained, in principle, by averaging the observations of the velocity of an element at a given (long) time after release from a given point in a large number of realizations or separate experiments) is equal to the average value, over a long time, of the velocity of an element in a single realization of the flow. Moreover, all the material elements in the pipe must have the same mean velocity, defined in either of the above two ways, since elements wander freely over the whole cross section and are not distinguishable from one another in any way. Therefore the probability average of the velocity of a material element is equal to the average of the velocities of a large number of different elements in any one realization, provided that the times at which these velocities are taken are spread over a large range.

The discharge velocity, on the other hand, is an average, over a large number of material elements in any one realization of the flow, of the velocity of an element as it arrives at the final cross section where the discharge is measured. Provided that the discharge velocity is determined from a measurement of the efflux over a long time, the efflux will be composed of large numbers of material elements with statistically independent velocities; and we see, from the argument of the preceding paragraph, that the discharge velocity, the probability average of the velocity of a material element and the velocity of a material element averaged a sufficiently long time are all equal.

It should be noted that the average of the instantaneous velocities of a large number of elements in any one realization is not necessarily equal to the probability average of the velocity of a particle; in fact, if surges occur in the pipe, these two averages will definitely not be equal. The essential requirement for equality of two averages of the velocity is that each of them should involve either a large number of realizations of the flow or a large range of times, these two ensembles being equivalent because $\mathbf{v}(t)$ is stationary with respect to t.

The above argument is very plausible but hardly rigorous, and the primary purpose of the experiments described in the following sections was to test the prediction that the discharge velocity and the average velocity of a single element over a sufficiently long time are equal. No previous investigation of the velocity of a material element seems to have been undertaken, although Allen and Taylor (1923) and Taylor (1954) have reported observations of the velocity with which a small quantity of injected salt solution travels down the pipe. They found that the velocity of the position of maximum concentration over a certain length of pipe was equal to the discharge velocity. This result is not as easily predicted as that given above, since their lengths of pipe were not so long that the standard deviation of the time of travel of a single material element injected with the salt solution was a negligible fraction of the total time. Evidently the result with salt solution depends, at least in part, on the fact that after a sufficiently long travel a large number of elements of fluid containing salt are far enough apart from each other to be regarded as wandering independently, so that the aggregate of their positions defines the probability distribution of the position of a single material element, which is gaussian with coincident positions of the maximum and the mean.

§ 3. DESCRIPTION OF EXPERIMENTAL APPARATUS

Water at about 40 ft head was supplied to a control valve from a large elevated tank, the level in which was kept constant to ± 1 in. by a pump. From the control valve the water passed to the experimental pipeline, 160 ft in length and composed of commercial steel tubing, 2 in. in nominal internal diameter, which had been previously in use. In the initial part of the pipeline an orifice-plate was inserted, and the air–water manometer indicating the pressure drop across the plate was set at the same reading throughout all the tests. Further downstream, the injection apparatus (figure 1) was mounted on a T-piece, and this was followed after a length of 8 ft by an easy 180° bend. From this point a straight horizontal length of 84 ft was installed, in which the measurements of time were made. The tubes were flanged, and before erection, the mean diameter of each length was determined by filling it with water and weighing the contents. The first timing slit A was placed 7 ft from the bend, and two others (B and C) were fitted at points 20·56 and 72·66 ft downstream from A. To ensure that the pipeline was always full, it terminated in an upward 270° loop, leading to a tank in which a sieve was placed to catch the particles. The water then fell over a weir into a measuring tank, by means of which the time required for a known volume of water to issue from the pipe (this time being of the order of 100 seconds) was measured at intervals. The discharge velocity U was $5{\cdot}01 \pm 0{\cdot}03$ ft sec^{-1}.

The particles were of diameters 0·1, 0·126, 0·2, 0·3, 0·4 and 0·6 in., and they were made of B.D.H. ‘Sira’ adhesive wax, which has a relative density 0·98. The material was weighed out to give a sphere of the required size and shaped into spherical form by rubbing between the hands. Grains of sand (and of lead shot for the largest size of particle) were then pressed into the particles until the measured terminal velocity in a large beaker of water was within the limits $\pm 0{\cdot}01U$.

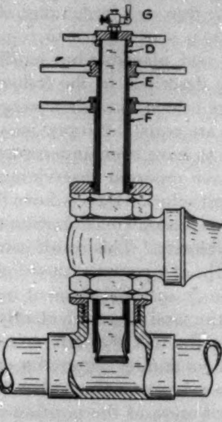


Figure 1. Injection apparatus.

The injection apparatus consisted of three brass tubes D, E and F, sliding within each other and provided with handles. Watertightness was maintained by Gaco rings. Tube F was prevented from moving by pressure from the isolating sluice valve, and E contained the particle. Tube D served to hold the particle close to the discharge ports in E, and it was blocked at its lower end by a plug in which two holes were drilled. To avoid the injection of air bubbles that would have actuated the timing instruments, the procedure of using the apparatus was as follows. After an injection, E was raised until the ports were only slightly uncovered, as shown in figure 1, this position being indicated by a distance piece inserted between the heads of E and F. The tap G was opened; and when water emerged, E was raised to close the ports completely. The tube D was pulled out and the particle dropped into E; D was then replaced, and D and E were lowered on to the distance piece. The tap G was shut when it was certain that the tubes were full of water; and after the distance piece had been removed, injection was made by forcing D and E downwards. The process could be repeated at intervals of 10 to 15 seconds. Since the pipeline was of considerable length, the protrusion of E into the stream did not sensibly reduce the discharge velocity. For the smallest three sizes of particle a smaller version of the same design was used.

Each detector unit of the timing apparatus, which was kindly devised by Mr. J. G. Yates, consisted of a 1-foot length of Perspex tube inserted in the pipeline. The curved surface was obscured with black paper except for a circumferential band 0·1 in. wide, and this cross section was watched by means of two 6-watt lamps, 90° apart and assisted by mirrors. The emergent light fell on a gas-filled photoelectric cell, and the pulse produced in the photocell current by the passage of a sphere was amplified and used to control the timer. Pulses at the rate of a hundred per second were provided by the 50 c/s mains and fed to two electronic gating circuits, which were opened by a pulse from unit A and closed by pulses from B and C. The timing pulses were counted on two three-decade 'Dekatron' cold-cathode circuits; thus the time intervals A–B and A–C were recorded directly in decimal form in units of 0·01 second.

§ 4. THE OBSERVATIONS OF PARTICLE VELOCITY

The time $T(x)$ taken by a single particle to travel between two points of observation distant x apart can be written as

$$T(x)=\overline{T}(x)+t', \qquad \ldots\ldots(1)$$

where $\overline{T}(x)$ is the probability average for particles of the kind considered. Then the average velocity of a particle over the distance x in a single realization of the flow is

$$u=\frac{x}{\overline{T}(x)+t'}=\frac{x}{\overline{T}(x)}\left\{1-\frac{t'}{\overline{T}(x)}+\left(\frac{t'}{\overline{T}(x)}\right)^2-\ldots\right\}. \qquad \ldots\ldots(2)$$

Taking the probability average over a large number of realizations, and denoting it by $U\,(\alpha)$ for spherical particles of radius αa where a is the pipe radius we have

$$U(\alpha)\simeq\frac{x}{\overline{T}(x)}\left\{1+\frac{\overline{t'^2}}{[\overline{T}(x)]^2}\right\}. \qquad \ldots\ldots(3)$$

The experiments, as analysed in §6, showed that $\overline{t'^2}/[\overline{T}(x)]^2\propto 2a/x$, with a constant of proportionality that varied between 0·14 and 0·04 according to the size of the particle. Hence the proportional error in calculating $U(\alpha)$ from

$$U(\alpha)=x/\overline{T}(x) \qquad \ldots\ldots(4)$$

is 0·28 a/x at most and is negligible in the experiments described here.

About 100 observations of the time of travel were made of each size of particle over the distances $x=20{\cdot}56$, $52{\cdot}10$ and $72{\cdot}66$ ft, a total of 570 observations on

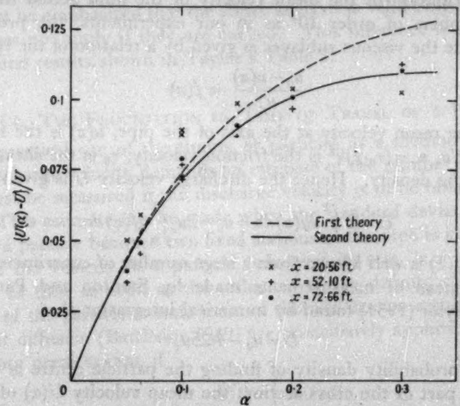


Figure 2. Effect of particle size upon particle mean velocity.

two different days being made of the 0·2 in. size. For each set of readings $T(x)$ was calculated, and then $U(\alpha)$ was determined from equation (4). Figure 2 shows the dependence of $\{U(\alpha)-U\}/U$ on α, the value of U being obtained from the measured discharge. It is clear that the experimental points lie about a curve passing through the origin, thus confirming the prediction that the discharge velocity and the probability average of the velocity of a single element of fluid are identical. A theory which accounts for the shape of the curve is developed in the next section.

§ 5. THE EFFECT OF FINITE PARTICLE SIZE

The theoretical result in § 2 is concerned with particles whose dimensions are vanishingly small, and some modification is necessary when particles of finite size are used to follow the motion. It was shown that a material element of fluid wanders freely over all parts of the region within the pipe, and that the probability of finding it in a specified small volume is independent of the position of that small volume. But when the radius of the pipe is a and the particle is a sphere of radius αa, the centre of the particle is free to move down the pipe only within a cylinder of radius $(1-\alpha)a$. It seems reasonable to suppose that the motion of the finite particle is such that the probability of finding its centre in a specified small volume is the same for all positions of that small volume inside a cylinder of radius $(1-\alpha)a$, and is zero for positions outside it. It is also necessary to make some assumption to relate the velocity of the particle to that of the fluid, and it will be supposed as a first approximation that the mean velocity of particles whose centres are at a specified radius r is equal to the mean velocity of the undisturbed fluid at the same radius. The particle ignores the annulus of thickness αa near the wall of the pipe where the fluid velocity is least, and we expect the mean velocity of the particle down the pipe to be greater than the discharge velocity.

The difference between these two velocities can be calculated from the observed distribution of the mean velocity of the fluid across the pipe. For Reynolds numbers of order 10^5 as in our experiments, the profile of mean velocity outside the viscous sublayer is given by a relation of the type

$$\frac{u_0-u(z)}{u_\tau}=f(z) \qquad \ldots\ldots(5)$$

where u_0 is the mean velocity at the axis of the pipe, $u(z)$ is the mean velocity at radius $z=r/a$, $u_\tau=(\tau_0/\rho)^{1/2}$ is the friction velocity, τ_0 is the shear stress at the wall and ρ is the density. Hence the discharge velocity U is given by

$$U=2\int_0^1 zu(z)\,dz=u_0-2u_\tau\int_0^1 zf(z)\,dz. \qquad \ldots\ldots(6)$$

The function $f(z)$ is well known from a large number of experiments. Inserting into (6) the mean of measurements made by Stanton and Pannell and by Nikuradse, Taylor (1954) found by numerical integration that

$$U=u_0-4{\cdot}25u_\tau. \qquad \ldots\ldots(7)$$

Since the probability density of finding the particle centre is uniform over the accessible part of the cross section, the mean velocity $U(\alpha)$ of a particle of radius αa can be found by averaging the same velocity profile over the circle of radius $z=1-\alpha$. Accordingly

$$U(\alpha)=\frac{2}{(1-\alpha)^2}\int_0^{1-\alpha} zu(z)\,dz=u_0-\frac{2u_\tau}{(1-\alpha)^2}\int_0^{1-\alpha} zf(z)\,dz \qquad \ldots\ldots(8)$$

and it follows from (7) that

$$\frac{U(\alpha)-U}{U}=\frac{4{\cdot}25-2(1-\alpha)^{-2}\int_0^{1-\alpha} zf(z)\,dz}{u_0/u_\tau-4{\cdot}25}. \qquad \ldots\ldots(9)$$

The experiments were performed at a Reynolds number $R=2aU/\nu=6{\cdot}6\times10^4$, where ν is the kinematic viscosity; and for this value of R Taylor's Fig. 2 for smooth pipes gives $U/u_\tau=20{\cdot}2$. From (7), $u_0/u_\tau=24{\cdot}5$, and (9) was evaluated as a function of α with values of the integral taken from Taylor's Table 1. The result is shown by the broken line in figure 2.

The assumption used in this simple theory, that the mean velocity of a particle at any point of the cross section is equal to the mean fluid velocity at the same radius as the particle centre, is a reasonable approximation only when the particle is so small that the mean velocity gradient does not vary appreciably over distances equal to the particle diameter. For a larger particle the assumption is in error as the curvature of the mean velocity profile is not zero. It is better to take the mean velocity of a particle as equal to the average value of $u(z)$ over the area of the cross section occupied by the particle. Since the curvature of $u(z)$ is negative, the mean velocity of a particle given by this assumption is less than that of the fluid at the same radial position as the particle centre. On this basis, with the same expression for $f(z)$ as before, the mean particle velocity was calculated for various values of α and z. The resulting profile for each value of α was then integrated over the cross section of radius $(1-\alpha)a$, yielding the full line in figure 2. At the limit $\alpha=1$, corresponding to a particle that fills the pipe, the first theory reduces to

$$\frac{U(\alpha)-U}{U}=\frac{u_0-U}{U}=4{\cdot}25\frac{u_\tau}{U}, \qquad \ldots\ldots(10)$$

whereas the second gives the correct result $U(\alpha)=U$.

It must be emphasized that the above analysis is valid only for straight pipes, for (5) does not apply if they are curved. This inference is supported by the experimental results shown in Taylor's Table 3.

§ 6. THE FLUCTUATION IN TIME OF TRAVEL OF A PARTICLE

Any practical use of the timing of particles as a method of determination of the discharge velocity demands an estimate of the number of particles whose times must be measured if the discharge velocity is to be obtained within given limits. This estimate can be made when the standard deviation of the time of travel of a particle between two fixed stations of the pipe is known.

Since the velocity of a particle, whatever its size, is a stationary random function of time as soon as the influence of the definite choice of the initial position of the particle is lost, some of the well-known results of the theory of turbulent diffusion (Batchelor 1949) are immediately applicable. In particular, that theory predicts that, if

$$X(t)=\int_{t_0}^{t} v(t)\,dt, \qquad \ldots\ldots(11)$$

where $v(t)$ is the component of the particle velocity in the direction of the axis of the pipe at time t, and t_0 is the instant of injection of the particle into the pipe, the probability distribution of $X(t)$ for a given value of t tends to the normal or gaussian form as $t-t_0 \to \infty$, and that in this asymptotic range

$$\frac{d}{dt}[\overline{X^2}(t)-\{\overline{X}(t)\}^2] \sim 2\int_0^\infty R(\xi)\,d\xi \qquad \ldots\ldots(12)$$

where

$$\overline{X}(t)=(t-t_0)\bar{v}, \qquad \ldots\ldots(13)$$

$$R(\xi)=\overline{[v(t)-\bar{v}][v(t+\xi)-\bar{v}]}. \qquad \ldots\ldots(14)$$

The integral in equation (12) depends on the size of the particle and on the distribution of velocity of fluid in the pipe. The latter is believed to be determined statistically by the radius of the pipe and the friction velocity u_τ everywhere outside the viscous sublayer; u_τ is itself determined by the Reynolds number R and the roughness of the wall. Hence, provided the existence of the viscous sublayer has no effect on the motion of the particle, we have on dimensional grounds that, when $t-t_0$ is large,

$$\frac{d}{dt}[\overline{X^2}(t)-\{\overline{X}(t)\}^2]=2au_\tau K'(\alpha) \qquad \ldots\ldots(15)$$

where $K'(\alpha)$ is an unknown function of the relative sizes of the particle and the pipe. $K'(\alpha)$, made dimensionless with the quantities $2a$ and u_τ, has the meaning of a longitudinal diffusivity for the spherical particles. At the Reynolds number of the experiments, the thickness of the viscous sublayer on a smooth wall is not more than about $0{\cdot}018a$ (Goldstein 1938), hence the above proviso was satisfied.

Equation (15) can be converted to an expression for the dispersion of the time $T(x)$ of travel over a given distance x, for

$$\frac{X(t)-\overline{X}(t)}{\overline{T}(x)-T(x)} \simeq \bar{v} = U(\alpha), \qquad \ldots\ldots(16)$$

when $$\overline{\{X(t)-(\overline{X}t)\}^2} \ll \overline{X}^2(t) \quad \text{and} \quad x=\overline{X}(t). \qquad \ldots\ldots(17)$$

Consequently

$$\frac{d}{dx}\overline{[T(x)-\overline{T}(x)]^2} \to \frac{2au_\tau}{U^3(\alpha)}K'(\alpha) \qquad \ldots\ldots(18)$$

as $x\to\infty$. Moreover, (16) shows that the probability distribution of $T(x)-\overline{T}(x)$ also tends to the normal form as $x\to\infty$. Equation (18) implies that

$$\frac{\overline{[T(x)-\overline{T}(x)]^2}}{\overline{T}^2(x)} \to \frac{2au_\tau}{xU(\alpha)}K'(\alpha)=\frac{2au_\tau}{xU}K(\alpha) \text{ say}, \qquad \ldots\ldots(19)$$

showing that the spread of times of travel is a diminishing fraction of the mean travel time as the length of pipe is increased. This fact corresponds to the known exact mathematical result that the probability average of the velocity of a particle is equal to the velocity of the particle in a single realization averaged over a sufficiently long time. Our experiments confirmed the relation (19), thus completing the demonstration that the discharge velocity and the velocity of a particle in a single realization averaged over a sufficiently long time are identical.

Our observations of the times of travel of particles enabled us to estimate the value of $K(\alpha)$ for six values of α. The values of $x\overline{[T(x)-\overline{T}(x)]^2}/\{a[\overline{T}(x)]^2\}$

did not vary systematically with x over the range $123 < x/2a < 436$, showing that the asymptotic value had been attained; the average values of $K(\alpha)$ obtained from measurements at three values of x in this range are shown in figure 3. The accuracy of these estimates is not high, and it is probable that more than 100 readings of the time of travel of a particle of given size are needed to establish a smooth curve. Making use of Reynolds' assumption that the lateral diffusivity of marked fluid is the same as that of momentum, Taylor calculated that $K(0) = 10{\cdot}1$. His measurements at $R \simeq 2 \times 10^4$ with salt solution as the marker of the fluid agreed with this value, which is considerably larger than the values shown in figure 3 for particles of finite size.

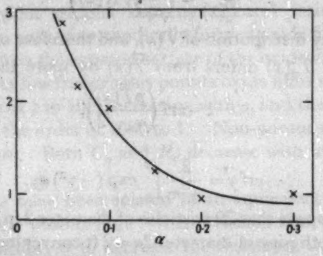


Figure 3. Effect of particle size upon the dispersion coefficient.

A large number of observations (570) of the time of travel of particles of diameter 0·2 in. ($\alpha = 0{\cdot}1$) was made in an endeavour to determine the probability distribution of the time of travel, and figure 4 shows histograms for $x/2a = 123$ and 436. Gaussian curves having standard deviations 0·101 and 0·189 second calculated from the observations are superimposed on the histograms, and the agreement is good, especially at the larger distance, in accordance with the theoretical prediction.

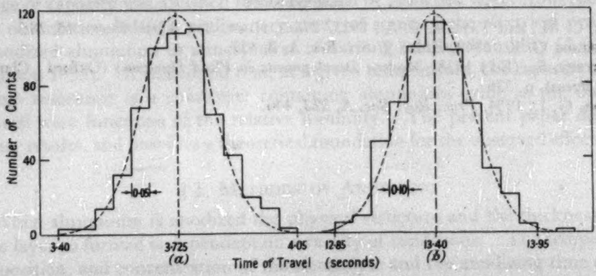


Figure 4. Histograms of the times of travel of 570 particles of diameter 0·2 in. over distances (*a*) $x = 20{\cdot}56$ ft, (*b*) $x = 72{\cdot}66$ ft, compared with gaussian distributions.

§ 7. THE PRACTICAL USE OF THE METHOD

In a practical application of the method as a means of obtaining the discharge velocity, the times of travel of a relatively small number, n say, of particles would be determined. To estimate the accuracy that can then be obtained, we make use of the elementary theory of statistics. If $T_n(x)$ denotes the average of n observations of the time of travel of a given particle over a given distance, it is known that the standard deviation of $T_n(x)$ about the true or probability average $\overline{T}(x)$ is given by

$$\sigma_n = \left[\overline{\{T_n(x) - \overline{T}(x)\}^2}\right]^{1/2} = \frac{\sigma}{n^{1/2}} \qquad \text{......(20)}$$

where σ is the standard deviation of single observations of the time of travel. Hence, making use of (19), we have

$$\frac{\sigma_n}{\overline{T}(x)} = \left(\frac{2au_\tau}{nxU} K(\alpha)\right)^{1/2}. \qquad \text{......(21)}$$

When the probability distribution of $T(x)$, and therefore of $T_n(x)$ also, is gaussian, the probability that $T_n(x)$ differs from $\overline{T}(x)$ by more than an amount $p\overline{T}(x)$ is

$$1 - \operatorname{erf}\left(\frac{p\overline{T}(x)}{\sigma_n\sqrt{2}}\right), \qquad \text{......(22)}$$

where

$$\operatorname{erf} y = \frac{2}{\sqrt{\pi}} \int_0^y \exp(-y^2)\, dy.$$

For example, suppose that 20 particles of diameter 2 in. are timed over a length $x = 300{\cdot}0$ ft of a smooth pipe of diameter $2a = 2$ ft conveying water at 15°C, and that the mean of the observations is found to be $T_n(x) = 30{\cdot}00$ seconds. For the purpose of estimating σ_n we take $U \simeq 10$ ft sec^{-1}, ignoring the difference between U and $U(\alpha)$. Then $\alpha = 0{\cdot}083$, $K(\alpha) = 1{\cdot}9$ from figure 3, $\nu = 1{\cdot}23 \times 10^{-5}$ ft^2 sec^{-1}, $R = 1{\cdot}63 \times 10^6$, $U/u_\tau = 27{\cdot}3$ from Taylor's Fig. 2, and $\sigma_n/\overline{T}(x) = 0{\cdot}0048$ from (21). The probability that $T_n(x)$ differs from $\overline{T}(x)$ by more than $0{\cdot}01\,\overline{T}(x)$ is thus

$$1 - \operatorname{erf} \frac{0{\cdot}01}{0{\cdot}0048\sqrt{2}} = 0{\cdot}038.$$

REFERENCES

Allen, C. M., and Taylor, E. A., 1923, *Trans. Amer. Soc. Mech. Engrs*, **45**, 285.

Batchelor, G. K., 1949, *Austr. J. Sci. Res.* A, **2**, 437.

Goldstein, S., (Ed.) 1938, *Modern Developments in Fluid Dynamics* (Oxford : Clarendon Press), p. 336.

Taylor, G. I., 1954, *Proc. Roy. Soc.* A, **223**, 446.

Some Electrical Effects of the Adsorption of Water Vapour by Anodized Aluminium

By A. C. JASON and J. L. WOOD

Torry Research Station, Food Investigation Organisation, Department of Scientific and Industrial Research, and Sir John Cass College, London E.C.3

MS. received 25*th November* 1954, *and in amended form* 18*th July* 1955

Abstract. The equivalent parallel capacity C_p and resistance R_p measured between the aluminium base of an anodically formed oxide film and a porous conducting layer deposited on the exposed surface of the oxide vary considerably with relative humidity ρ. At low frequencies porous oxide films show a range of variations of C_p of the order of 1 to 100, increasing with ρ, and the corresponding range of variation of R_p is of the order of 1000 to 1. Non-porous oxide films show very much smaller variations. Both C_p and R_p decrease with increasing frequency f for porous films.

These observations have been related to an equivalent circuit based on the physical structure of the porous film. It is assumed that the principal effect of increasing relative humidity is to decrease the resistance down the length of the pore sides. A quantitative comparison confirms that this accounts both for the very large variations of C_p and R_p with ρ and for their variation with f.

---

§ 1. Introduction

Ansbacher (1949) showed that the capacity of a condenser containing anodized aluminium oxide as the dielectric material was very considerably reduced when placed in a container which was then evacuated. The change of capacity was ascribed to the removal of adsorbed water from the oxide. This observation led to a preliminary study of some of the electrical properties of anodized aluminium in atmospheres of varying relative humidity (Ansbacher and Jason 1953). It was found that, at a given temperature, the capacity and the leakage resistance of a condenser containing aluminium oxide as the dielectric material were functions of the relative humidity. The present paper describes further results, and provides a theoretical foundation for the observed effects.

§ 2. Methods of Anodizing

When aluminium is anodized the physical structure and the thickness of the oxide layer so formed is dependent on a variety of conditions. The temperature, composition, and concentration of the electrolyte and the anodizing time may be varied independently, and either the formation voltage or the current density may be controlled in any desired manner (von Zauscher 1935, Napier and Westwood 1950, Bahn and Böttger 1953, Charlesby 1953). Thus, depending on the selection of these variables, a great variety of thicknesses, porosities and structures may be produced. In the work to be described, attention was principally directed to the four types of oxide layers anodized by the methods given in table 1.

Table 1. Some Dielectric Properties of various Anodic Films measured in the electrolyte and when dry

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Amm. borate (sat. sol.)	40	45	—	180	0·45	3·32 (Boreskov)	381000	56000	6·8	6470	612	10·6
Chromic acid 50 g CrO_3/l.	40	90	—	100	12·9	2·28	312000	1790	174	1300	2620000	$4{\cdot}95\times10^{-4}$
	(Bengough–Stuart process)											
Oxalic acid 3·78 g $(COOH)_2.2H_2O$/l.	40	90	—	30	10·3	2·81	194000	1140	170	4840	7320000	$6{\cdot}60\times10^{-4}$
Sulphuric acid 17·5% v/v	30	—	50	30	31·2	2·06	3700000	615	6020	209	6920000	$3{\cdot}02\times10^{-5}$

(1)–(5) Anodizing details : (1) electrolyte and concentration, (2) temperature (°C), (3) maximum formation voltage, (4) current density (mA cm^{-2}), (5) time (min); (6) oxide thickness $d(\mu)$; (7) oxide density (gm cm^{-3}); (8) capacity in electrolyte C_e(pF); (9) capacity when dry C_0(pF); (10) C_e/C_0; (11) resistance in electrolyte $R_e(\Omega)$; (12) resistance when dry $R_0(\Omega)$; (13) R_e/R_0.

The capacity and resistance measured at 1000 c/s of oxide layers prepared in each of these electrolytes were measured *in situ* in the anodizing bath immediately after formation of the oxide film. The electrolyte served as one conductor and the aluminium sample, in the form of a rod, as the other. The samples (4 cm long × 0·158 cm diameter) were then washed in tap water for a few minutes, rinsed in distilled water for several hours, and a porous conductive electrode of 'Aquadag' was deposited on the surfaces. After they had stood over phosphorus pentoxide for twenty-four hours, the capacity and resistance of each of these samples were again measured. These measurements are given in table 1. The film formed in ammonium borate is known to be non-porous (Hass and Kehler 1941, Hass 1949). The thickness of this film was estimated from the loss of weight of the sample after chemically removing the oxide, and an assumed oxide density of 3·32 gm cm^{-3} (Boreskov *et al.* 1952). The thicknesses of the porous films formed in the acid electrolyte were determined by direct measurements on other samples prepared simultaneously and their bulk densities were derived from their weight and dimensions. Table 1 shows that the capacity and the resistance of the non-porous sample formed in ammonium borate suffered comparatively little change when transferred from the electrolyte to a dry atmosphere, relative to the changes which are apparent with porous samples formed in the acid electrolytes. The strong dependence of these changes on the porosity is reflected in the effects of water vapour on the samples, described in the following section.

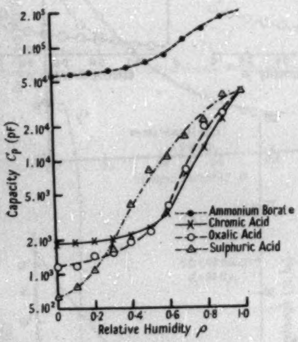


Figure 1. The influence of the type of oxide layer on the variation of capacity with relative humidity at 20°c.

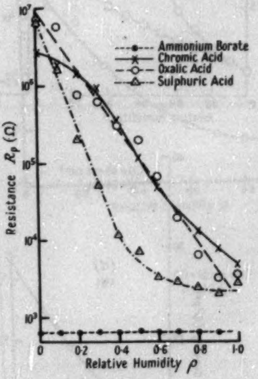


Figure 2. The influence of the type of oxide layer on the variation of resistance with relative humidity at 20°c.

This remarkable effect of humidity, giving rise to apparent dielectric constants of 1000 or more, which thus appears to be associated with the porosity of the layers, has been investigated in detail. The effect is explained on the basis of the following observations.

§ 3. INFLUENCE OF ELECTROLYTE AND FILM THICKNESS

3.1.

Measurements were made of the capacity and resistance at 1000 c/s of various samples placed in an atmosphere in which the relative humidity was varied while the temperature remained constant at 20°c. The samples were always formed on rods and coated with an outer electrode of 'Aquadag'.

Isotherms relating the equivalent parallel capacity C_p and resistance R_p to relative humidity ρ for the samples listed in table 1 are shown in figures 1 and 2. The difference between the non-porous sample prepared in ammonium borate and the porous samples prepared in the acid electrolytes is very apparent. It will be noticed that on approaching saturation C_p and R_p do not become equal to

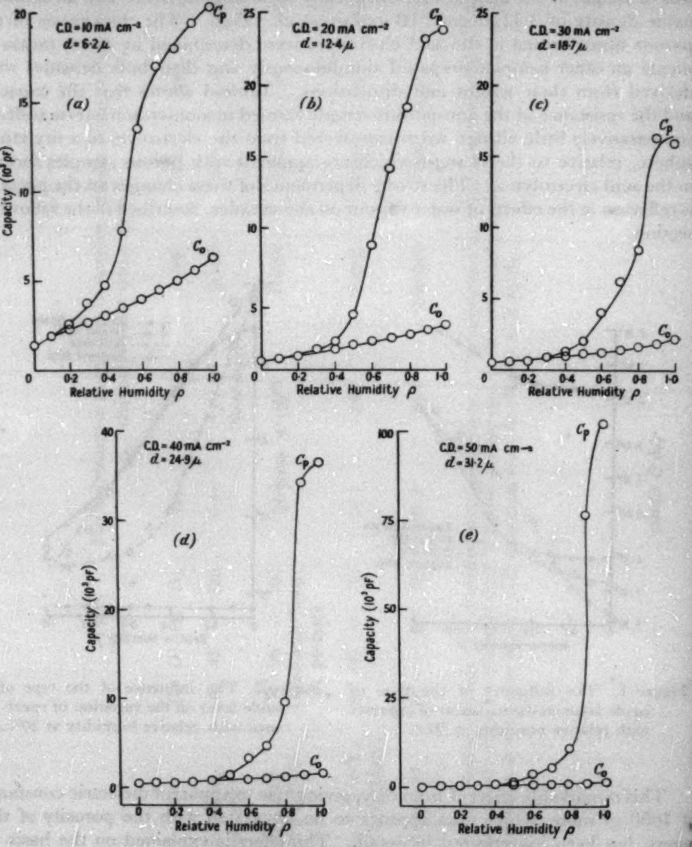


Figure 3. Effect of film thickness d on capacity C_p.

C_e and R_e (the capacity and resistance measured in the electrolyte) of the same samples. This is probably due to rapid changes occurring in the oxide structure on the cessation of anodizing.

3.2.

Samples of aluminium were anodized at constant current density in sulphuric acid for 30 minutes. Their thicknesses were proportional to the current density: 0·62 ± 0·02 micron per mA cm^{-2} current density. The capacity and resistance isotherms at 20°c corresponding to layers of various thickness are shown in figures 3 and 4.

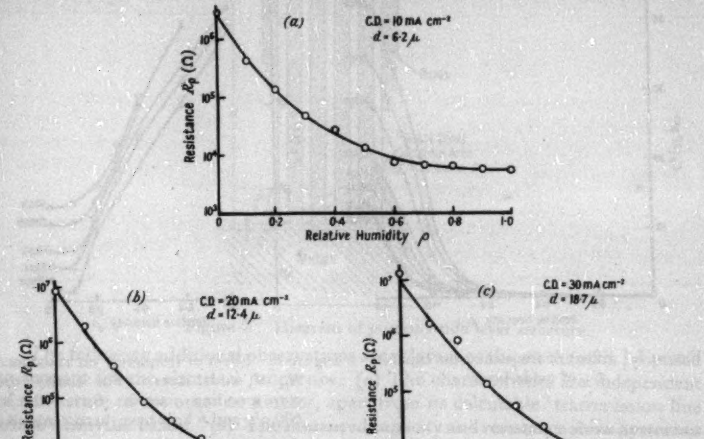


Figure 4. Effect of film thickness *d* on resistance R_p.

§ 4. DEPENDENCE OF CAPACITY AND RESISTANCE ON FREQUENCY

The element used to investigate the effect of bridge frequency on the measured values C_p and R_p was anodized in sulphuric acid as before, but the outer conducting layer was formed of evaporated aluminium. On this was wound a helix of thin aluminium wire, reducing the resistance along the outer conductor to less than 1 ohm. This is always negligible compared with the transverse impedance of the oxide layer, so simplifying the theoretical treatment of the observations.

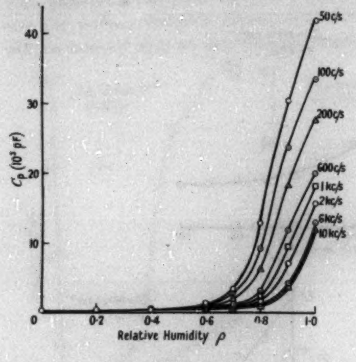


Figure 5. Effect of frequency on capacity C_p of oxide layer.

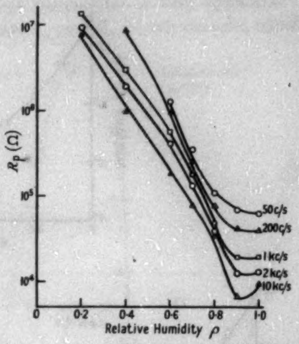


Figure 6. Effect of frequency on resistance R_p of oxide layer. (For clarity, the corresponding curves for 100 c/s, 600 c/s and 6 kc/s have been omitted.)

Capacity and resistance measurements, made on this element in the range of frequency from 50 c/s to 10 kc/s at selected relative humidities from 0 to 1·0 at 20°c are presented in figures 5 and 6. These show that both capacity and resistance decrease with increasing frequency at all relative humidities. An interpretation of the effects shown in figures 5 and 6 is given in the following section.

§ 5. THEORY

Electron microscopic examination of the structure of these porous films (to be published) indicates that the pores are uniformly distributed over the entire surface, parallel, approximately perpendicular to the surface and run through almost the entire oxide layer. There remains a thin layer of oxide between the bases of the pores and the metal. A diagram of this pore structure is shown in figure 7.

The overall thickness d of the oxide layer is determined by the charge passed in the course of the entire anodizing process, in accordance with Faraday's law and the rate of solution of alumina in the electrolyte (Napier and Westwood 1950, Huber 1948, Keller *et al.* 1953). For anodization at 30°c in 17·5% sulphuric acid at 50 mA cm^{-2} for 30 minutes, for example, d is $31 \cdot 2 \pm 1 \cdot 2 \times 10^{-4}$ cm. The two other dimensions which principally determine the capacity and resistance isotherms of porous oxide layers are those of the pore base thickness *b*, and of the pore area *S* near the base. With the elements prepared in the above way, *b* is of the order of 120 Å and the pore base area is approximately 10% of the superficial surface area. Since the pore is possibly not exactly a right cylinder near the base the values of *b* and the cross section are only approximations, giving the order of magnitude of the pore base dimensions.

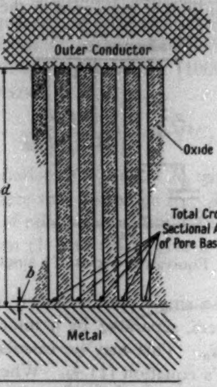


Figure 7. Diagram of porous oxide layer structure.

The following additional observations are relevant to the mechanism proposed to account for the electrical properties: (*a*) The characteristics are independent of the nature of the outer conductor, apart from its calculable 'transmission line effect' (Smythe 1950). (*b*) The measured capacity and resistance show hysteresis after exposure to relative humidities not less than 0·85. The extent and direction of the hysteresis is similar to that shown in the absorption of water by alumina. This suggests that the equilibrium values of the electrical properties are due to water adsorption, and hysteresis to water absorption. (*c*) The capacity at relative humidities not greater than 0·3 is equal to the capacity of a condenser distance *d* between plates, and of a mixed water–air–alumina dielectric calculated from the addition equation (Böttcher 1952, eqn (11.4), p. 415). This is C_0 shown in figure 3. (*d*) In an atmosphere of relative humidity approaching 100%, or in liquid water, the apparent dielectric constant on the basis of a thickness *d* rises to a value of more than 1000.

This increase is far greater than can be attributed to saturation by water, or to any geometrical disposition of a mixed water–alumina dielectric (Böttcher 1952, § 64). Measurements of the dielectric constant of water in the adsorbed phase are scanty, but the observations of McIntosh, Rideal and Snelgrove (1951) for the capacity of water adsorbed on silica give a value of the dielectric constant which is not greater than that of the bulk liquid. In any case, the extremely high observed value is too large to be attributed to the modification of the alignment of the adsorbed water molecules on adsorption.

We suggest that with increasing humidity, electrolytic conduction down the pore surface occurs, so that the outer conducting layer then follows the aluminium oxide contours. The capacity is then largely determined by the contribution between the pore base and the metal. Since the distance *b* is very small, this capacity will be high. The dimensions of the pore base are quantitatively consistent with these high capacities observed at saturation, on the assumption that the dielectric is alumina. The individual variations from sample to sample of the saturation capacity are probably due to variations in the pore base thickness. A theory which accounts quantitatively on this basis for the observed electrical behaviour at all relative humidities is presented in the following paragraphs.

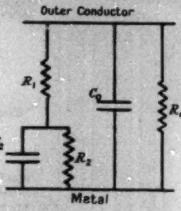


Figure 8. Equivalent circuit of a single pore.

The equivalent circuit of a single pore is shown in figure 8. The capacity between two conducting layers, separated by a mixed air, alumina and water dielectric of thickness *d* is C_0. This may be calculated from the quantity of water adsorbed, by Böttcher's equation (11.4). When the relative humidity is not greater than 0·3, C_0 accounts for nearly the entire measured capacity. The leakage resistance through the full distance *d* of the oxide layer is R_0. The remainder of the equivalent circuit is due to the specific contribution from the pore. The area at the base of a pore, where *b* is small, provides almost the entire contribution to C_2, the pore base capacity. The resistance R_1 is that along the pore side between the outer conducting layer and C_2. This resistance is strongly dependent on the quantity of water adsorbed on the pore side surface. R_2 is the leakage resistance through the thickness *b* of alumina at the pore base.

For a single pore, the parallel capacity equivalent to the circuit in figure 8 is, at frequency *f*,

$$C_p = \frac{C_2 R_2^2}{(R_1+R_2)^2 + 4\pi^2 f^2 R_1^2 R_2^2 C_2^2} + C_0 \qquad \ldots\ldots(1)$$

and the equivalent parallel resistance is given by

$$\frac{1}{R_p} = \frac{R_1 + R_2 + 4\pi^2 f^2 R_1 R_2^2 C_2^2}{(R_1+R_2)^2 + 4\pi^2 f^2 R_1^2 R_2^2 C_2^2} + \frac{1}{R_0}. \qquad \ldots\ldots(2)$$

For a set of identical pores, represented by a number of the above circuit elements in parallel, two formally identical equations result, the symbols now referring to the entire oxide layer.

The implication of these equations is that at low humidities at which the pore side resistance R_1 becomes large, $C_p - C_0$ tends to zero, i.e. the contribution from the 'pore effect' is very small. As the relative humidity rises toward saturation, R_1 falls to a small value, so that C_p tends to $C_0 + C_2$ and R_p tends to R_2. The term $1/R_0$ is negligible when *f* is small compared with the relaxation times of the mixed dielectric of C_0.

The data of figures 3 and 4 showing the dependence of C_p and R_p on the thickness d provide a quantitative confirmation of equations (1) and (2). At relative humidities not greater than 0·3 the measured capacity C_p is inversely proportional to the thickness d, and proportional to the quantity of adsorbed water. These relationships are those of C_0, and the entire measured capacity C_p is due to this term. As the relative humidity rises above 0·3 towards saturation the contribution of C_0 to C_p is strongly dominated by the pore contribution. C_0 is calculated from the quantity of adsorbed water (the values are plotted in figure 3) and subtracted from C_p obtained experimentally, to give $C_p - C_0$. At the frequency used in these measurements (1000 c/s), $1/R_0$ is negligible, so that (1) and (2) combine to give

$$4\pi^2 f^2 (C_p - C_0) + \frac{1}{R_p{}^2 (C_p - C_0)} = 4\pi^2 f^2 C_2 + \frac{1}{R_2{}^2 C_2} = A. \quad \ldots\ldots(3)$$

We shall assume that the variation of C_2 and of R_2 with relative humidity is small compared with that of the pore side resistance R_1; A, the right-hand side of (3), then becomes independent of relative humidity. In other words, the parameter R_1 has been eliminated from (1) and (2) to give a direct relationship between R_p and $C_p - C_0$.

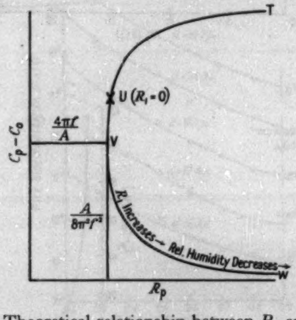


Figure 9. Theoretical relationship between R_p and $C_p - C_0$.

The form of equation (3) is shown in figure 9. The curve TUVW represents the equation, the parameter R_1 varying along this curve. The point U, where $R_1 = 0$, lies in the branch VT when $2\pi f C_2 R_2 > 1$. In this case, values of $(R_p, C_p - C_0)$ such that $4\pi^2 f^2 (C_p - C_0) > 1/R_p{}^2 (C_p - C_0)$ lie on VU, and the explicit solution of (3) is

$$C_p - C_0 = \frac{A}{8\pi^2 f^2} + \tfrac{1}{2}\left[\left(\frac{A}{4\pi^2 f^2}\right)^2 - \frac{1}{\pi^2 f^2 R_p{}^2}\right]^{1/2}. \quad \ldots\ldots(4)$$

In all other cases the negative sign of the root is to be taken. In the data given in this section the negative sign condition always holds, i.e. R_p decreases as $C_p - C_0$ increases.

For each of the five porous layers (figures 3 and 4) the experimental value of $C_p - C_0$ is plotted against R_p (figure 10). The values of $C_p - C_0$ calculated by (3) from R_p are also shown. The agreement is most satisfactory. The values of the constant A for each curve are shown in table 2.

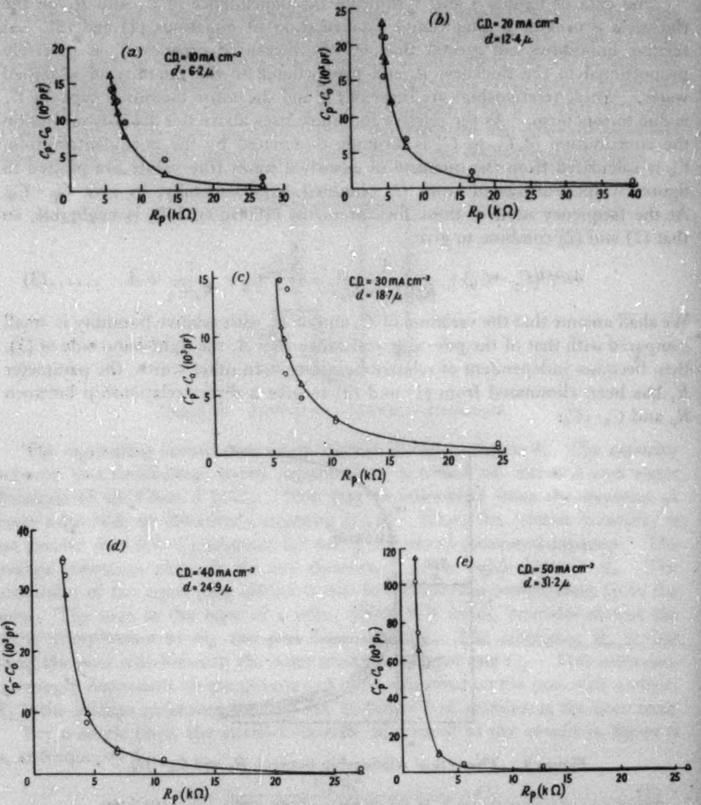


Figure 10. Values of $C_p - C_0$ plotted as a function of R_p for various thicknesses d of the oxide layer. Circles represent experimental values and triangles theoretical values

Table 2. Pore-base Capacity and Resistance

(1)	(2)	(3)	(4)	(5)	(6)
6·2	10	2·52	19 000	5·5	$1{\cdot}04 \times 10^{-4}$
12·4	20	3·04	25 000	4·38	$1{\cdot}10 \times 10^{-4}$
18·7	30	3·2	15 800	5·0	$0{\cdot}79 \times 10^{-4}$
24·9	40	5·5	40 000	2·53	$1{\cdot}01 \times 10^{-4}$
31·2	50	9·2	118 000	1·35	$1{\cdot}59 \times 10^{-4}$

(1) Oxide thickness $d(\mu)$; (2) forming current density (mA cm⁻²); (3) A; (4) C_2(pF); (5) R_2(kΩ); (6) C_2R_2(ΩF).

Since $A=4\pi^2f^2C_2+1/R_2{}^2C_2$, an approximate value of A may be obtained from the measured C_p-C_0 and R_p of a layer as the humidity tends to saturation (see equations (1) and (2)). However, this method has two objections: (*a*) the entire (R_p, C_p-C_0) curve is then based on one experimental point; (*b*) even at saturation R_1 may not be zero. The values of A chosen are therefore the mean values obtained by substituting C_p-C_0 and R_p in (3). The values of A would be given by the values of C_2 and R_2 shown in table 2. (The probable error in the estimate of these is $\pm 10\%$, except for $d=18{\cdot}7\mu$ when it is $\pm 20\%$.) The product C_2R_2 is characteristic of the nature of the dielectric at the pore base and independent of its shape. It should therefore be constant. Table 2 shows that the variation of C_2R_2 with d is much less than that of either C_2 or R_2 individually. The calculated values of R_1, the pore side resistance, at each of the points, are shown in figure 11. R_1 shows a marked decrease with increasing relative humidity. In general, R_1 also increases with d, that is, as the length of the pore side increases. Due to the sensitivity of these calculated values of R_1 to small errors in measurement, the values given are approximate, particularly at relative humidities of 80% or more.

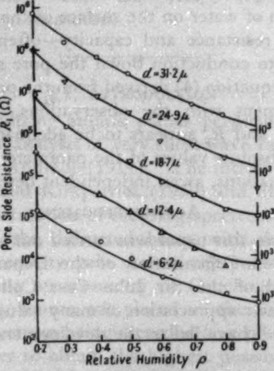


Figure 11. Calculated values of pore side resistance R_1 as function of ρ and d.

5.1. *Effect of Frequency*

Equations (1) and (2) indicate that the values of C_p and R_p at a given relative humidity will vary with the frequency used in measurements.

At relative humidities approaching 1·0 and frequencies above 1 kc/s the resistance R_L along the entire length of an 'Aquadag' outer conducting layer becomes comparable with the impedance across the element. Although this 'transmission' line effect can be calculated (Smythe 1950), it was found simpler to reduce R_L as previously described. The curves relating R_p to C_p for this element at various frequencies (figure 12) follow equation (3), the value of the constant A varying with the frequency. This implies that C_2 and/or R_2 is not independent of f.

It has been noted that when $2\pi fC_2R_2>1$, the curve of figure 9 will have a VU branch. Thus at humidities approaching saturation (when the parameter $R_1\to 0$) and at higher frequencies, we may expect R_p to increase while C_p-C_0

increases, on approaching saturation. This minimum in the value of R_p is found near saturation in all the curves for frequencies greater than 1 kc/s (figure 12).

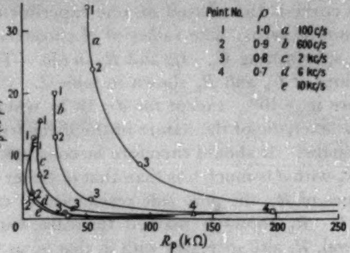


Figure 12. Variation of C_p and R_p with ρ at various frequencies.

§ 6. CONCLUSION

The resistance and capacity of porous anodized aluminium oxide layers are related to the adsorption of water on the surface of the pores. The remarkably large variations of the resistance and capacity—often observed to be several hundred-fold—are due to conduction down the pore side to the large capacity at the pore base. The equation (4) derived from the proposed equivalent circuit gives reasonable agreement with the observations. However, although the relationship between C_p and R_p appears to be adequately accounted for by the equivalent circuit, the absolute values of the parameters, such as R_1, C_2 and R_2, which enter into the equations, await independent determination.

ACKNOWLEDGMENTS

The work described in this paper was carried out as part of the programme of the Food Investigation Organisation of the Department of Scientific and Industrial Research and of the Sir John Cass College. The authors wish particularly to express their appreciation of many valuable suggestions made by Mr. W. W. Foster, which have led to the development of the theory presented in this paper.

The paper is published by permission of Her Majesty's Stationery Office.

REFERENCES

ANSBACHER, F., 1949, *Ph.D. Thesis*, University of London.
ANSBACHER, F., and JASON, A. C., 1953, *Nature, Lond.*, **171**, 177.
BAHN, R., and BÖTTGER, O., 1953, *Z. Phys.*, **135**, 376.
BORESKOV, G. K., BORISOVA, M. S., DZISKO, V. A. and KRASNOPOLSKAYA, V. N., 1952, *J. Phys. Chem., Moscow*, **26**, 492.
BÖTTCHER, C. J. F., 1952, *Theory of Electric Polarisation*, Elsevier.
CHARLESBY, A., 1953, *Proc. Phys. Soc.* B, **66**, 317.
HASS, G., 1949, *J. Opt. Soc. Amer.*, **39**, 532.
HASS, G., and KEHLER, H., 1941, *Kolloid Z.*, **97**, 27.
HUBER, K., 1948, *J. Colloid Sci.*, **3**, 197.
KELLER, F., HUNTER, M. S., and ROBINSON, D. L., 1953, *J. Electrochem. Soc.*, **100**, 413.
MCINTOSH, R., RIDEAL, E. K., and SNELGROVE, J. A., 1951, *Proc. Roy. Soc.* A, **208**, 292.
NAPIER, D. H., and WESTWOOD, J. V., 1950, *Met. Ind.*, **76**, 123.
SMYTHE, W. R., 1950, *Static and Dynamic Electricity*, 2nd edn (London and New York; McGraw-Hill), p. 377.
VON ZAUSCHER, H. 1935., *Ann Phys., Lpz.*, **23**, 597.

A Precision Curved Crystal Gamma-ray Spectrometer

By N. RYDE and B. ANDERSSON
Department of Physics, Chalmers University of Technology, Göteborg, Sweden

MS. received 25th April 1955; *and in final form* 13*th June* 1955

Abstract. A curved crystal spectrometer of transmission type with a radius of curvature of one metre has been constructed, utilizing the (310) planes of quartz. The effective part of the crystal plate has the dimensions 10 mm × 20 mm × 1 mm and subtends 0·00020 steradian at the focus. In order to ensure a high degree of stability the spectrometer is mounted on a strong U-beam. The spectrum is scanned by moving a slit in front of the radiative source along the Rowland circle, the direction of the diffracted beam being fixed in space. The γ-rays are detected by means of a NaI(Tl) scintillation crystal combined with a single channel pulse height discriminator. The wavelengths and energies of gamma rays from ^{198}Au, ^{76}As, ^{122}Sb, ^{175}Yb, ^{177}Lu, ^{186}Re, ^{188}Re, ^{192}Ir, and ^{194}Ir, obtained by pile activation have been measured.

---

§ 1. Introduction

The principle of utilizing atomic planes perpendicular to the surface of a crystal for spectral analysis of very short wave radiation was first applied by Rutherford and Andrade (1914). The introduction of curved crystals proposed by Du Mond and Kirkpatrick (1930) and first realized by Cauchois (1932) has led to the construction of focusing spectrometers in which the rapid decrease of intensity of the diffracted beam with decreasing wavelength is compensated by the luminosity gain due to the curved crystal. In the first transmission type focusing spectrometer of Cauchois, intended for study of x-rays, an extended x-ray source was placed on the convex face of the curved crystal, the diffraction spectrum being recorded on a photographic film along the Rowland circle. In order to obtain the highest possible luminosity, necessary for measurements on γ-rays, Du Mond (1947) made use of the reverse configuration placing the radioactive source in such a position that by moving it along the circumference of the Rowland circle the intensity variations of the diffracted beam could be recorded by means of a multi-cellular Geiger–Müller counter. The beam transmitted directly through the crystal was absorbed by a collimator of tapered lead sheets. Du Mond constructed at Pasadena a spectrometer for precision measurements of γ-rays according to these principles with a crystal bent to a radius of 2 metres. The γ-ray sources consisted of thin activated metal foils, placed with the surface perpendicular or nearly perpendicular to the Rowland circle. With this instrument Du Mond *et al.* (1948–1953) carried out measurements on the energies of γ-rays from ^{198}Au, ^{131}I, ^{60}Co, ^{192}Ir, ^{182}Ta, ^{183}Ta, Ra, Th, Rn, ^{187}W, ^{137}Cs, and from the annihilation of positrons from ^{64}Cu. In the last case a thicker irradiated copper plate was applied behind a slit. By installing certain auxiliary optical equipment Muller *et al.* (1952) were able to make measurements on the instrument which permitted corrections for some mechanical flexures in parts of it. These corrections, together with some improvements in luminosity and sensitivity, made it possible to increase the precision of the wavelength measurements. A preliminary report of the design of the spectrometer used in the present investigation was given by Andersson (1954). In this spectrometer the γ-ray detector is a thallium-activated NaI scintillation crystal. A detector of this type was installed at about the same time in the Pasadena spectrometer, replacing the multi-cellular Geiger–Müller counter formerly used.

§ 2. GEOMETRICAL OPTICS OF THE SPECTROMETER

The 1 mm thick quartz crystal K (figures 1, 2) was bent in a holder to a cylindrical surface with a radius of curvature of 1015 mm. The Rowland circle has the line AB as diameter where A is the mid-point of the neutral surface of the bent lamina and B is the centre of curvature of this surface's mid-section. Thus the atomic planes perpendicular to the crystal surface intersect, if produced, at B. After diffraction all rays emanating from a 'real focus' R on the circle seem to come from the virtual focus V. This holds exactly for the central part of the crystal; at the extremities of the crystal, however, the fact that the neutral axis does not coincide with the Rowland circle gives rise to an aberration. This aberration has been studied by Cauchois (1932). It results in a relative line broadening

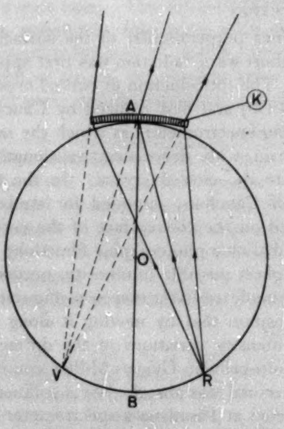


Figure 1. Principal geometry of focusing by means of a curved crystal (transmission case).

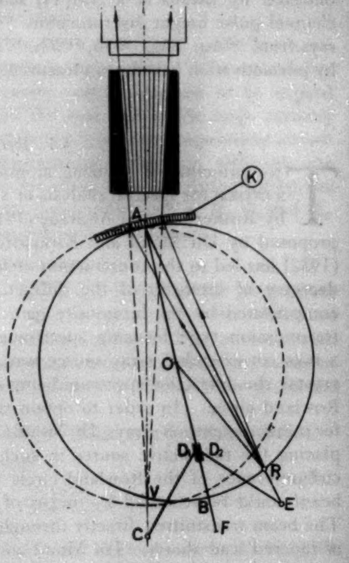


Figure 2. Principal geometry of the curved crystal spectrometer.

$$\frac{\Delta\lambda}{\lambda} \simeq \frac{\cos\theta(1-\cos\alpha)}{\cos(\alpha+\theta)}$$

where θ is the Bragg angle and 2α the angle subtended by the mid-section of the effective part of the crystal at the point B on the Rowland circle. For the present spectrometer $\Delta\lambda/\lambda$ amounts to 0·00005.

§ 3. KINEMATICAL DESIGN OF SPECTROMETER

The spectrometer is mounted on an iron U-beam 200 cm long and 30 cm wide, placed on a concrete pillar erected directly on the rock and isolated from the remaining part of the institute building. Figure 2 is a schematic drawing of the spectrometer. The detector is stationary and the beam striking it fixed in space. The design must then fulfil the following conditions: (*a*) the real focus R has always to be on the Rowland circle, (*b*) the virtual focus V must always be on the line of symmetry through the middle of the crystal and of the detector.

The crystal holder may be rotated around a vertical axis through A, the middle of the crystal, by the arm AF, consisting of a steel rod 2 cm in diameter fastened to the crystal holder. AB is the radius of curvature of the crystal and O the mid-point of AB. O is the centre of the Rowland circle and the arm OR, of the same length as OA and revolving around O, forces the point R, the intersection of the slit jaws in front of the radiating source, always to follow the Rowland circle, thus satisfying condition (*a*). AR is realized by a U-beam, 130 cm long and 6·5 cm wide, which may be rotated around a vertical axis coinciding with that of the crystal holder. This U-beam supports the radioactive source and the slit in front of it. The slit holder slides along the U-beam. In order to fulfil the second condition the angles VAB and RAB must always be equal, which is ensured by the two equal arms CD_1 and ED_2, the former pivoted around a vertical axis C fixed to the main steel beam, the latter around a vertical axis E attached to the beam AR. The arms CD_1 and ED_2 are connected at D_1D_2, and their common vertical axis of rotation D_1 D_2 slides along the arm AF. Thus, when the beam AR is turned around the axis A through the crystal, the Rowland circle is turned through half the angle, and the virtual focus V always remains on the spectrometer's axis of symmetry as demanded by condition (*b*).

The movement of the U-beam AR about the axis A is controlled by a cross-slide arrangement and micrometer screw. The beam is extended about 28 cm behind the source holder and near the end is supported by a vertical steel cylinder I_1 (figure 3) attached to a carriage movable in the longitudinal direction of the beam, the guide for this carriage being placed on another carriage movable in a direction perpendicular to the spectrometer's axis of symmetry. The U-beam AR (figure 2) is pivoted around an axis R coinciding with the axis of the steel cylinder. The turning of this beam, by which the whole spectrum is scanned, is thus achieved by moving the lower carriage, which is done by means of a precision screw with a pitch of 0·5 mm. The displacement of the carriage may be read on a cylindrical scale to 0·0001 mm. The upper and lower carriage, as well as the source holder, all move on straight ball-bearings. All rotational movements are supported by circular ball-bearings.

§ 4. THE CRYSTAL HOLDER

The quartz crystal, 1 mm thick and with the (310) planes (spacing 1177·63 X.U. at 18°C) normal to the flat face of the lamina, is clamped in the crystal holder with the (310) planes vertical. The screws S_1 and S_2 (figure 4, Plate) press it between steel blocks B_1 and B_2, which are very carefully turned and ground to a circular cylindrical profile. In order to allow for adjustment the crystal holder (figures 4, 5, 6) is mounted in a gimbals arrangement. It is held in the cradle C_1 by means of the screws S_3, S_4. This cradle can revolve around a horizontal axis A_1A_2 through the middle of the crystal, the movement being controlled by the screw S_5. The axis has bearings in another cradle C_2, revolving around an axis perpendicular to A_1A_2 and in the direction of the rays. The rotation around this axis is controlled by means of the screw S_6. The whole arrangement is mounted on a steel plate P_1 supported by a vertical cylindrical steel rod I_2 and may be rotated around the axis of the rod passing through the middle of the crystal. This axle passes through a ball-bearing attached to the steel beam U_1 supporting the instrument, and its motion is controlled by the steel bar AF (figure 2) which

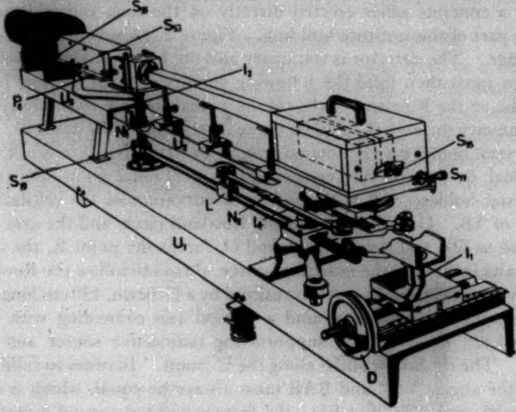


Figure 3. Perspective drawing of the γ-ray spectrometer.

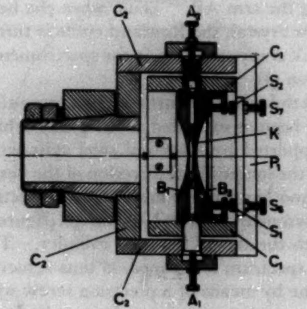


Figure 5. Horizontal section through the crystal holder.

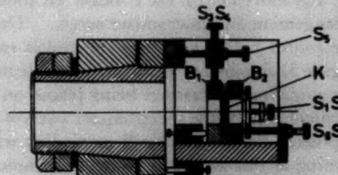


Figure 6. Vertical longitudinal section through the crystal holder.

is fixed to the axle I_2 by means of the screw S_{19} (figure 3). Another ball-bearing permits the U-beam AR to rotate around the same axis. The axle I_2 is divided into two parts connected by a nut N_1, the upper and lower parts of which have right-hand and left-hand threads respectively. By turning the nut the crystal holder may thus be raised or lowered. The lower part of the axle I_2 is slotted, as is also the lower end of the upper part of the axle. A steel pin in this slot prevents the crystal holder from rotating when it is raised or lowered by turning the nut N_1.

§ 5. THE SOURCE HOLDER AND SLIT

The source holder is screwed to a carriage which is attached to the arm AR (figure 2) and slides along the U-beam U_2 on a straight roller bearing (figure 3). Directly connected to the carriage is a short steel cylinder I_3 (figure 7) and above this a circular iron plate P_2. A second equally large iron plate P_3 above the first rests on three adjusting screws S_8, S_9 and S_{10}, and supports the peripheral part of a large cylindrical ball-bearing with vertical axis. The iron plate P_4 (18 cm × 26 cm × 0·7 cm) is fixed to the central part of the ball-bearing, and underneath it a lead plate P_5 (18 cm × 26 cm × 2 cm) is fastened. The iron plate forms the bottom of the source holder box. The walls of the box are made of lead and are 4 cm thick. The lid is of lead 3 cm thick and is removable for insertion of the radiative source. All adjustments are made with screws outside the source holder box. It may be rotated about the axis of the ball-bearing by means of the micrometer screw S_{11} (figures 3, 7, 8), which is fixed to the arm M, this in its turn being connected to the outer part of the ball-bearing. The source holder and the slit may be moved by means of the screw S_{15} (figure 8) in a longitudinal direction and by means of the screw S_{13} (figures 7, 8) in a transverse direction. The source material itself is contained in aluminium boxes, the inner walls of which are inclined to each other. The space between the walls is filled with the metal to be investigated, either in the oxide or metallic form, and after irradiation in the Harwell pile (BEPO) at a high neutron flux, the aluminium container is

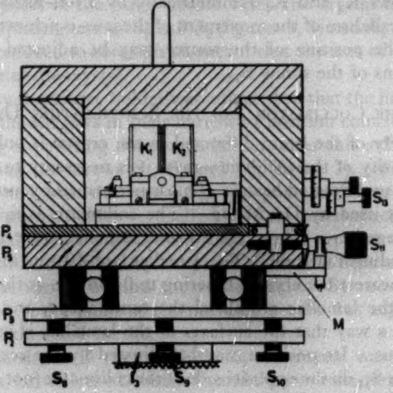


Figure 7. Vertical cross section through the source holder.

pressed between the iron blocks K_1 and K_2 (figures 7, 8, 9, Plate). The radiating source thus becomes wedge-shaped with the narrower part towards the slit. The total length of the container is 6 cm. Owing to the limited size of the irradiation units in the pile, the container is divided into three parts, each 2 cm long, 0·4 cm wide and 3·7 cm high. Depending upon the strengths to which the sample may be activated, one, two or three sections of the container may be used. The effective sample height is 1·6 cm. The slit jaws K_3 and K_4 (figure 8) are made of steel blocks into which lead plates are fitted. The surfaces of these plates were carefully made plane. The angle between them is 1·1°, this being the angle subtended by the crystal at R (figures 1, 2, 8), the point of intersection of the slit jaws. The slit width may be varied by means of the screw S_{14} (figure 8) which moves the slit jaws K_3 and K_4 symmetrically by a left-hand and a right-hand thread. The parallelism of the movement of the jaws is achieved to a high degree of precision. The position of the source may be adjusted in the transverse direction by means of the screw S_{12}.

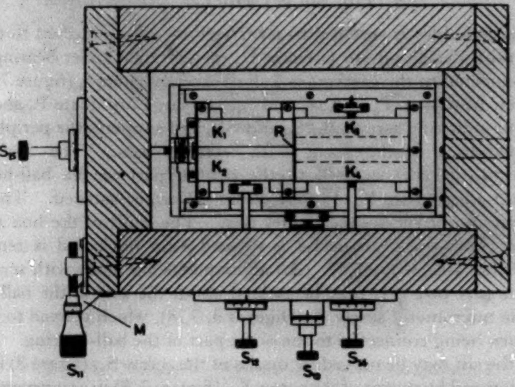


Figure 8. Horizontal section through the source holder.

§ 6. THE COLLIMATOR AND THE SCINTILLATION DETECTOR

As the intensity of the beam diffracted in the crystal is only a small fraction of the total intensity of the incident γ-rays, it is necessary to ensure very good absorption of the undiffracted beam. This is performed by means of a laminated collimator as first used by Du Mond. The collimator is made of 40 cm steel blocks screwed together. There are 20 steel lamellae, very accurately worked in a surface grinding machine. The lamellae are 10 mm high and 0·559 mm thick at the end nearest the crystal, tapering to 0·756 mm at the other end. The spaces between the lamellae are all of the same shape, and the collimator is arranged in such a way that the surfaces of the lamellae, if produced, intersect at the virtual focus. Its position may be adjusted by the screws S_{16} (figure 3) and another screw S_{17} on the opposite side of the collimator (not seen in the figure). The collimator rests on a steel plate P_6 which is levelled by means of three screws, one of which (S_{18}) brings the lamellar planes parallel to the atomic planes of the crystal. A U-beam U_3 supports the collimator and detector tube, which is housed in a lead cover 4–5 cm thick. The detector first used was an EMI 5311 photomultiplier tube, and later a Du Mont 6292 photomultiplier tube in conjunction with a cylindrical NaI(Tl) crystal (diameter $1\frac{3}{4}$ in., height $\frac{1}{2}$ in.). The scintillation pulses were recorded with a single channel differential discriminator of conventional design.

§ 7. METHODS OF ADJUSTING THE SPECTROMETER

The upper surface of the large U-beam U_1 is very accurately ground under the source to a distance of 48 cm from the end. This was also done with the surface under the support of the crystal holder, and the two ground parts of the beam surface made parallel with the utmost possible care. These surfaces were first made horizontal by means of a spirit level. The crystal was then clamped in the crystal holder to a radius of curvature of about 100 cm and the accuracy of the bending tested by inspecting the image of an illuminated vertical slit obtained by reflection at the curved crystal surface, the slit being parallel to the generators of the curved surface. The clamping screws were adjusted until the image of the slit became sharp. The radius of curvature was then measured as accurately as possible, and a further test made by photographing an x-ray diffraction spectrum in a special arrangement. The Ag K_2 lines were investigated on each side of the line of symmetry. When the curvature of the crystal had in this way been found to be correct the crystal holder was placed in the spectrometer and adjusted so that the horizontal axis of revolution A_1A_2 (figures 4, 5) passed through the centre of the crystal. Next, the middle of the crystal, the slit and the collimator lamellae were adjusted to the same level with the aid of a cathetometer a few metres distant. In order to adjust the crystal so that its centre was on the vertical axis of revolution of the crystal table, the screw S_{19} (figure 3) fixing the latter to the bar AF (figure 2) was loosened. An upright needle was placed on the crystal holder vertically above the centre of the crystal. Another needle on a tripod support was brought into position above the first, with the needle point in the centre of the circle described by the point of the other needle when the crystal table was rotated. By means of the adjusting screws S_1 and S_2 (figures 4, 5) and A_1 and A_2 the radius of this circle was reduced to zero, which meant that the axis of revolution passes through the centre of the crystal.

In a similar way the source holder was adjusted so that the line of intersection of the surfaces of the slit jaws in their narrowest position coincided with the axis of the ball-bearing supporting the source holder, after this axis had been made vertical by adjusting the screws S_8, S_9 and S_{10} (figure 7). The source holder support was loosened with the screw S_{11}, the point of an upright needle was placed against the source holder vertically above the intersection of the slit jaws, and another vertical needle placed on a fixed tripod support with the point at the centre of the circle described by the point of the first needle when the source holder support was turned around the vertical axis. By means of the screws S_{13} and S_{15} (figure 8) the point of the movable needle was adjusted to fall just under the point of the fixed needle.

By turning the drum D (figure 3) the spectrometer was brought into its symmetry position, i.e. with the U-beam U_2 and the two rods corresponding to OR and AF in figure 2 above each other. A strip of black paper with a Λ-shaped hole in it was placed at the slit with its surface perpendicular to the longitudinal direction of the spectrometer and with the point of the hole at the middle of the slit. The latter was illuminated from behind and the image of the point made to coincide with the point itself by adjusting the screw S_5 (figures 4, 6). In this way the generators of the bent crystal surface were made parallel to the axis of revolution of the crystal holder. The distance AO (figure 2) was then made equal to half the value of the radius of curvature of the bent crystal, determined in the way described above. This was accomplished by loosening the junction L (figure 3) and regulating the distance between the axis of revolution of the crystal table and that of the arm I_4 (figure 3) at L, representing the radius of the Rowland circle, using gauge blocks (made by C. E. Johansson). The centre of the Rowland circle thus having been fixed the junction L was tightened and the distance between the axes of revolution of the crystal table and the slit holder made equal to the radius of curvature of the bent crystal by turning the nut N_2 on the arm I_4 connecting the axis through the centre of the Rowland circle with the carriage supporting the slit holder. Gauge blocks were also used for this adjustment. The radius of the Rowland circle was then equal to half the radius of curvature of the bent crystal. When this had been done, the crystal table was turned until the image of the slit obtained by reflection at the crystal fell on the slit itself, whereupon the crystal table was fixed with the screw S_{19} to the bar AF (figure 2). The collimator was then coarsely adjusted by an optical method so that the lamellar surfaces when produced intersected at the slit.

Further adjustments were performed with a radioactive substance in the source holder. The spectrometer position for a strong γ-line having been calculated, pulses were counted for a series of settings around the calculated position. The same procedure was followed for the position on the other side of the line of symmetry. By adjusting the screw S_{20} (figure 4) the crystal holder could be turned around a horizontal axis through the middle of the crystal and its centre of curvature. The half-widths of the curves of counting rate against spectrometer setting for a series of such crystal positions were determined, and the minimum value gave the position where the crystal's reflecting planes were parallel to the slit. The curves obtained in this way also enabled the symmetry position of the spectrometer to be calculated. The spectrometer was brought to this setting, a counting-rate meter was switched in instead of the scaler and mechanical register, and the collimator adjusted to maximum deflection of the counting-rate meter. Again using the latter as detector, the source holder support was turned around the vertical axis of rotation until the position for maximum deflection of the counting-rate meter was obtained. The adjustments of collimator and source holder were repeated until no further improvements could be attained, after which the source holder support was fixed. For all these adjustments the narrowest possible slit was used.

The resolving power of the spectrometer depends upon the design geometry as well as upon imperfections of the crystal and faults made at the adjustment. The aberration arising because the neutral axis of the crystal does not coincide with the Rowland circle (see § 2) and that due to crystal diffraction have only a small influence on the resolving power. To the effect of the imperfections of the crystal and the errors of the adjustments must be added that of the height of the crystal and the width of the slit. For intensity reasons these quantities cannot be reduced below certain limits. The resolving power attained in practice seems to be determined mainly by the extensions of the effective source. Defining the resolving power as the percentage width of a gamma-ray line at half maximum height, we obtain from the experimentally determined line profiles a resolving power of about 1% for the energy region in question.

§ 8. MEASUREMENTS

After the radiative source had been placed in position in the source holder and its position adjusted by means of the screw S_{12} to give maximum radiation intensity at the detector, using the counting-rate meter with the spectrometer in zero wavelength position, the slit was set to a suitable width. Usually the slit widths varied between 0·10 and 0·15 mm. The discriminator was then set to respond to pulse heights corresponding to the gamma radiation to be investigated, the spectrometer still being in the symmetry position. By choosing a channel width as small as possible the background was reduced to a minimum. The counting-rate meter was replaced by the scaler and mechanical register, and the line profiles on opposite sides of the symmetry position were taken by counting the pulses for periods of 2–15 minutes, depending on the gamma radiation in question, over a series of different settings. The main screw of the spectrometer was always turned in the same direction in moving from one setting to another. From the line profiles the exact position of the two spectrometer settings for the γ-ray line could be determined. After several runs for each line the mean positions were computed, and the distance between the two positions of the carriage measured by means of gauge blocks. During the runs the temperature was kept constant to within $\pm 0{\cdot}3°$ of 22°C. The errors given are maximum deviations from the mean values.

Gold ^{198}Au

The 0·412 MeV radiation from ^{198}Au has earlier been subject to very thorough measurements. The first precision determination was performed by Du Mond, Lind and Watson (1948) with the Pasadena crystal spectrometer, and resulted in the energy value $E_{\mathrm{Au}} = 0{\cdot}4112 \pm 0{\cdot}0001$ MeV. Later Lind and Hedgran (1951) calculated $E_{\mathrm{Au}} = 0{\cdot}4111 \pm 0{\cdot}0004$ MeV from measurements with a high precision two-directional focusing β-spectrometer as designed by Svartholm and Siegbahn. Using the external conversion method they compared the 0·412 and 1·31 MeV energies of radiation from ^{198}Au and ^{60}Co, referring the E_{Au} value to the E_{Co} value $1{\cdot}3318 \pm 0{\cdot}001$ MeV, determined by Lind, Brown and Du Mond (1949) with the Pasadena crystal spectrometer. Absolute measurements on the ^{198}Au 0·412 MeV line with the β-spectrometer combining measurement of K and L photo-electron lines from an external converter according to a method proposed by Siegbahn yielded the energy value $E_{\mathrm{Au}} = 0{\cdot}41155 \pm 0{\cdot}0004$ MeV, showing a discrepancy with respect to the crystal spectrometer value. This discrepancy was still more marked when Hedgran and Lind later (1952), using the β-spectrometer mentioned above, calculated E_{Au} firstly from comparison between the ThL line and the K internal conversion line of ^{198}Au, secondly from a similar comparison between the ThL line and the L_{III}(U) line for ^{198}Au, based on the $H\rho$-value for the ThL line which had been determined by Lindström from β-spectrometer measurements. The result obtained by Hedgran and Lind was

$$E_{\mathrm{Au}} = 0{\cdot}41175 \pm 0{\cdot}00010 \text{ MeV}.$$

After having carried out a new calibration of the Pasadena crystal spectrometer with the aid of auxiliary optical equipment Muller, Hoyt, Klein and Du Mond (1952) made a new measurement of the ^{198}Au 0·412 Mev line, using a source of 5 curies initial activity. This strong source permitted measurement of second, and third-order reflections too. The weighted mean value of ten runs in the first order, one run in the second order, and two runs in the third order obtained by Muller *et al.* was $E_{Au}=0{\cdot}411770\pm0{\cdot}000036$ Mev. This value is, within the limits of error stated, in excellent agreement with the β-spectrometer determination. As these two determinations, both carried out with the utmost care in totally different ways, agree so very closely, we must have confidence in the results. If the two E_{Au} values are assigned weights proportional to the reciprocal squares of the probable errors, we obtain $E_{Au}=0{\cdot}41176_7\pm0{\cdot}00004_8$ Mev. Of all γ-ray lines the strong gold line is at present the one with the most carefully determined energy, and therefore very suitable as a standard line, to which other measurements may be referred. As gold may be easily and quite rapidly brought up to high activities in a pile, the gold line is very convenient as a reference line for measurements with crystal spectrometers. For these reasons the gold line has been chosen as a standard for all energy determination measurements of the present paper, using the weighted mean value of the best earlier measurements as given above.

Several gold samples have been investigated. The best measurements were obtained with a wedge-shaped foil of 2 g weight brought up to an initial activity of about 10 curies, and the calibration was based upon these measurements. Six runs were made for the line profiles in the first order on each side of the zero wavelength position. As it appeared that the ascending and descending sides of the line profiles were accurately linear, except for the highest and lowest parts, the exact position of the γ-ray line was determined by the intersection of the lines for the two sides of the peak, these lines being deduced from the observed points by the method of least squares. In this way the difference between the spectrometer settings for the ^{198}Au 0·412 Mev line on opposite sides of the zero wavelength position was found to be $6{\cdot}3105\pm0{\cdot}0024$ cm. From this value the distance between the axes of rotation of the crystal holder and of the supporting cylinder I_1 is computed to be 123·434 cm, using the wavelength value 30·105 x.u. for the gold line.

Arsenic ^{76}As

The radiation from 26·8-hour ^{76}As was first studied by Miller and Curtiss (1946) with a magnetic lens spectrometer. They observed one strong γ-ray line of 0·57 Mev energy, one gamma line of medium intensity of 1·25 Mev energy and two very weak lines of higher energies. With a similar instrument Siegbahn (1947) determined the energy of the strong γ-ray line to 0·548 Mev. Using a helical focusing magnetic spectrometer Wu, Havens and Rainwater (1948) obtained for the strong line the energy value 0·557 Mev. For the same line Marty-Wollman (1951), using a semicircular magnetic spectrometer, found the energy to be 0·567 Mev, and with a magnetic lens spectrometer Hubert (1951) obtained the energy $0{\cdot}555\pm0{\cdot}002$ Mev.

A sample of metallic arsenic irradiated in the pile and having an initial activity of 1·5c at the beginning of the measurements was investigated with the crystal spectrometer. The mean of two runs gave the wavelength $22{\cdot}116\pm0{\cdot}010$ x.u. and the energy $0{\cdot}5605\pm0{\cdot}0003$ Mev for the strong line.

Antimony ^{122}Sb

The radiation from 66-hour ^{122}Sb was first accurately measured by Cook and Langer (1948) who found a γ-ray line with the energy 0·568 Mev. Glaubman and Metzger (1952) observed another line at 0·68 Mev, in cascade with the former one. Recently Cork *et al.* (1954) reinvestigated the ^{122}Sb radiation. They found the strong line at 0·5660 Mev and five fainter ones at 0·0952, 0·553, 0·616, 0·647, and 0·694 Mev, all of which they ascribe to ^{122}Sb.

A pile-irradiated sample of antimony metal with an initial activity of 200 mc at the beginning of the measurement was analysed in the crystal spectrometer. The mean of three runs gave for the strong line the wavelength $21·982 \pm 0·007$ x.u. and the energy $0·56393 \pm 0·00019$ Mev.

Ytterbium ^{175}Yb

102-hour ^{175}Yb has earlier been investigated by Cork *et al.* (1950) who with a β-spectrometer made measurements on four γ-ray lines, the strongest one being found to have the energy 0·396 Mev. Only this line could be accurately measured in the crystal spectrometer, using a sample containing more than 99·5% Yb_2O_3 with an initial activity of 0·5c. As a mean value of two runs we obtained for the strongest ^{175}Yb line the wavelength $31·358 \pm 0·009$ x.u. and the energy value $0·39532 \pm 0·00012$ Mev.

Lutecium ^{177}Lu

Accurate measurements of the radiations from 7-day ^{177}Lu were first made by Douglas (1949), and by Cork, Keller and Stoddard (1949) using β-spectrometers. The latter made measurements on two strong lines and found the energies to be 0·1131 and 0·2086 Mev. For these lines Douglas obtained the energy values

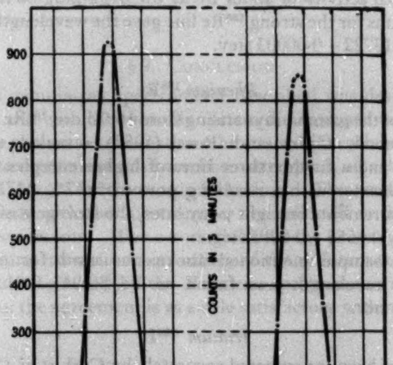


Figure 10. Profile of the 0·208 Mev line of ^{177}Lu at both sides of the symmetry position.

0·1122 ± 0·0006 and 0·2063 ± 0·0010 Mev. He showed that these gamma-rays are emitted in cascade, and was also able to determine the energy of the weak cross-over line, 0·3137 ± 0·0015 Mev. Since the present investigation was completed measurements by Marmier and Boehm (1955) with the Pasadena crystal spectrometer have been published, giving energy data for the three lines mentioned above and in addition for two very weak lines in cascade, the latter having the energies 0·071644 ± 0·000020 Mev and 0·2500 ± 0·0005 Mev. For the two strong lines Marmier and Boehm found the energies to be 0·112965 ± 0·000020 Mev and 0·208362 ± 0·000020 Mev. The energy obtained for the cross-over line was 0·32136 ± 0·00010 Mev.

A sample containing 98% Lu_2O_3 was irradiated in the pile. The initial activity was 1c. Four runs for the 0·113 Mev line and five runs for the 0·208 Mev line were performed. Other radiation could not be detected with certainty. The wavelengths and energies of the lines were found to be 109·627 ± 0·024 x.u., corresponding to 0·11308 ± 0·00003 Mev and 59·638 ± 0·010 x.u., corresponding to 0·20786 ± 0·00004 Mev. From these energies the energy and the wavelength of the cross-over line is 0·32094 ± 0·00007 Mev, corresponding to 38·625 ± 0·009 x.u.

Rhenium ^{186}Re

The gamma radiation from 92·8-hour ^{186}Re has been investigated by Beach, Peacock and Wilkinson (1949), using a semicircular β-spectrometer. They found two γ-ray lines, the stronger one with the energy 0·138 Mev. Metzger and Hill (1951), with a β-spectrometer, found the energy of this line to be 0·137 Mev, and observed two further lines. Applying a similar technique Steffen (1951) obtained the value 0·136 ± 0·001 Mev for the strong line.

A sample of rhenium metal containing more than 99·9% Re was irradiated in the pile to an initial activity of about 2 c at the beginning of the measurements. The mean of six runs for the strong ^{186}Re line gave the wavelength 90·34 ± 0·02 x.u. and the energy 0·13722 ± 0·00003 Mev.

Rhenium ^{188}Re

Investigation of the gamma rays arising from 16·9-hour ^{188}Re with a β-spectrometer by Richmond, Grant and Rose (1952) revealed a strong line at 0·152 ± 0·001 Mev and a further three lines of higher energies. Using a double focusing β-spectrometer with a resolving power of 0·7% McMuller and Johns (1953) made measurements on eight γ-ray lines, the strongest of which was found to have the energy 0·1553 ± 0·0005 Mev.

The rhenium sample mentioned above was used for measurements on this line. The wavelength was found to be 80·04 ± 0·03 x.u., the energy 0·15487 ± 0·00005 Mev.

Iridium ^{192}Ir

74-day ^{192}Ir has been investigated accurately by Cork *et al.* (1951), who used a magnetic β-spectrometer to make measurements on 13 lines belonging to ^{192}Ir, the strongest ones having the energies 0·6112, 0·6037, 0·4674, 0·3161, 0·3077, and 0·2949 Mev. Bashilov *et al.* (1952) were able to identify 15 ^{192}Ir lines and found the energies of the strongest ones to be 0·613, 0·605, 0·468, 0·317, 0·309 and 0·296 Mev. Muller and collaborators (1952) measured very accurately with the Pasadena crystal spectrometer energies and intensities for eleven ^{192}Ir lines, For the four strongest lines they found the energies 0·467984 ± 0·000061, 0·316462 ± 0·000034, 0·308454 ± 0·000033, and 0·295942 ± 0·000031 MeV, and for two fainter cross-over lines 0·61287 ± 0·00061 and 0·60453 ± 0·00059 MeV.

An iridium sample was irradiated in the pile for a week. The activity at the beginning of the measurements was about 700 mc, making it possible to determine the energy of the four strongest lines of ^{192}Ir and one line of ^{194}Ir (see below). For the 0·468 MeV line we obtained as a mean of six runs, the wavelength 26·491 ± 0·007 X.U. and the energy 0·46795 ± 0·00012 MeV. For the strongest line four runs gave the wavelength 39·165 ± 0·005 X.U. and the energy 0·31652 ± 0·00004 MeV. Three runs for the 0·308 MeV line gave the wavelength 40·162 ± 0·007 X.U. and the energy 0·30866 ± 0·00006 MeV. The 0·316 and 0·308 MeV lines were well separated, as were also the 0·308 and 0·295 MeV lines. For the latter the wavelength was found to be 41·885 ± 0·015 X.U. and the energy 0·29596 ± 0·00010 MeV. As pointed out by Cork *et al.* and by Muller and collaborators, the 0·295 and 0·308 MeV as well as the 0·295 and the 0·316 MeV lines are in cascade, the energies of the cross-over lines may thus be calculated as 0·60462 ± 0·00016 MeV and 0·61248 ± 0·00014 MeV. The corresponding wavelengths are 20·503 ± 0·005 X.U. and 20·240 ± 0·005 X.U. The results of the measurements are generally within the limits of error in agreement with those of Muller and collaborators.

Iridium ^{194}Ir

19-hour ^{194}Ir has been investigated by Cork *et al.* (1951) who found the energy of the strongest line to be 0·3275 MeV, using a magnetic β-spectrometer. Baker, Bleuler, and Steffen (1952) observed two further lines, and found the energy of the strong line to be 0·326 MeV. In the crystal spectrometer measurements could only be made on this line, and as the mean of two runs we obtained the wavelength 37·671 ± 0·010 X.U. and the energy 0·32907 ± 0·00008 MeV.

§ 9. CONCLUSION

The precision gamma-ray spectrometer described was designed with special rigid foundations and with the movable system dimensioned to avoid mechanical flexures. For the present determinations the attainable resolving power was about 1%. No absolute determinations have been attempted, all values being referred to the $E_{\mathrm{Au}} = 0{\cdot}41176_7$ MeV for the strong ^{198}Au line. The results of the measurements hitherto made with the spectrometer are collected in the table, including some earlier measurements. The errors given for the present measurements are maximum deviations from the mean values of the different runs. Where comparison with results achieved by the Pasadena spectrometer is possible, i.e. for ^{192}Ir and for ^{177}Lu, the agreement is as a rule satisfactory within the limits of error stated.

ACKNOWLEDGMENTS

The authors are indebted to several of the staff of the Department of Physics, Chalmers University of Technology, for valuable assistance. Our thanks are in particular due to Mr. Bengt Lanne for helpful co-operation in the construction of the spectrometer, and to Mr. S. Nilsson for assistance during measurements. The work was assisted by a grant from the Swedish Atomic Energy Commission.

		Present measurements		Earlier measurements			
		Wavelength (x. u.)	Energy (Mev)	Energy (Mev)			
				Siegbahn (1947)	Wu *et al.* (1948)	Marty-Wollman (1951)	Hubert (1951)
^{76}As		22·116 ± 0·010	0·5605 ± 0·0003	0·548	0·557	0·567	0·555 ± 0·002
				Cook & Langer (1948)	Cork *et al.* (1954)		
^{122}Sb		21·982 ± 0·007	0·56393 ± 0·00019	0·568	0·5660		
				Cork *et al.* (1950)			
^{175}Yb		31·358 ± 0·009	0·39532 ± 0·00012	0·396			
				Cork *et al.* (1949)	Douglas (1949)	Marmier & Boehm (1955)	
^{177}Lu		59·638 ± 0·010	0·20786 ± 0·00004	0·2086	0·2063 ± 0·0010	0·208362 ± 0·000020	
		109·627 ± 0·024	0·11308 ± 0·00003	0·1131	0·1122 ± 0·0006	0·112965 ± 0·000020	
	calc.	38·625 ± 0·009	0·32094 ± 0·00007		0·3137 ± 0·0015	0·32136 ± 0·00010	
				Beach *et al.* (1949)	Metzger & Hill (1951)	Steffen (1951)	Anton'eva *et al.* (1953)
^{186}Re		90·34 ± 0·02	0·13722 ± 0·00003	0·138	0·137	0·136 ± 0·001	0·1383 ± 0·0005
				Richmond *et al.* (1952)	McMuller & Johns (1953)		
^{188}Re		80·04 ± 0·03	0·15487 ± 0·00005	0·152 ± 0·001	0·1553 ± 0·0005		
				Cork *et al.* (1951)	Bashilov *et al.* (1952)	Muller *et al.* (1952)	
^{192}Ir	calc.	20·240 ± 0·005	0·61248 ± 0·00014	0·6112	0·613	0·61287 ± 0·00061	
	calc.	20·503 ± 0·005	0·60462 ± 0·00016	0·6037	0·605	0·60453 ± 0·00059	
		26·491 ± 0·007	0·46795 ± 0·00012	0·4674	0·468	0·467984 ± 0·000061	
		39·165 ± 0·005	0·31652 ± 0·00004	0·3161	0·317	0·316462 ± 0·000034	
		40·162 ± 0·007	0·30866 ± 0·00006	0·3077	0·309	0·308454 ± 0·000033	
		41·885 ± 0·015	0·29596 ± 0·00010	0·2949	0·296	0·295942 ± 0·000031	
				Cork *et al.* (1951)	Baker *et al.* (1952)		
^{194}Ir		37·671 ± 0·010	0·32907 ± 0·00008	0·3275	0·326		

References

Andersson, B., 1954, *Ark. Fys.*, **7,** 73.

Anton'eva, N. M., Bashilov, A. A., Dzhelepov, B. S., and Chervinskaya, L. S., 1953, *Izv. Akad. Nauk SSSR*, Ser. Fiz. **17**, 507.

Baker, D. T., Bleuler, E., and Steffen, R. M., 1952, *Perdue University Progress Report II* (June).

Bashilov, A. A., Anton'eva, N. M., and Dzhelepov, B. S., 1952, *Izv. Akad. Nauk SSSR*, Ser. Fiz. **16**, 264.

Beach, L. A., Peacock, C. L., and Wilkinson, R. G., 1949, *Phys. Rev.*, **76,** 1585.

Cauchois, Y., 1932, *C. R. Acad. Sci., Paris*, **194,** 362, 1479.

Cook, C. S., and Langer, L. M., 1948, *Phys. Rev.*, **73,** 1149.

Cork, J. M., Keller, H. B., Rutledge, W. C., and Stoddard, A. E., 1950, *Phys. Rev.*, **78,** 95.

Cork, J. M., Keller, H. B., and Stoddard, A. E., 1949, *Phys. Rev.*, **76,** 986.

Cork, J. M., LeBlanc, J. M., Stoddard, A. E., Childs, W. J., Branyon, C. E., and Martin, D. W., 1951, *Phys. Rev.*, **82,** 258.

Cork, J. M., Brice, M. K., Hickman, G. D., and Schmid, L. C., 1954, *Phys. Rev.*, **93,** 1059.

Douglas, D. G., 1949, *Phys. Rev.*, **75,** 1960.

DuMond, J. W. M., 1947, *Rev. Sci. Instrum.*, **18,** 626.

DuMond, J. W. M., Hoyt, H. C., Marmier, P. E., and Murray, J. J., 1953, *Phys. Rev.*, **92,** 202.

DuMond, J. W. M., and Kirkpatrick, H. A., 1930, *Rev. Sci. Instrum.*, **1,** 88.

DuMond, J. W. M., Lind, D. A., and Watson, B. B., 1948, *Phys. Rev.*, **73,** 1392; 1949, *Ibid.*, **75,** 1226.

Glaubman, M. J., and Metzger, F. R., 1952, *Phys. Rev.*, **87,** 203.

Hedgran, A., and Lind, D. A., 1952, *Ark. Fys.*, **5,** 177.

Hubert, P., 1951, *J. Phys. Radium*, **12,** 823.

Lind, D. A., Brown, J. R., Klein, D., Muller, D. E., and DuMond, J. W. M., 1949, *Phys. Rev.*, **75,** 1544.

Lind, D. A., Brown, J. R., and DuMond, J. W. M., 1949, *Phys. Rev.*, **76,** 1838.

Lind, D. A., and Hedgran, A., 1951, *M. Siegbahn, Commem. Vol.*, p. 283; also 1952, *Ark. Fys.*, **5,** 29.

Marmier, P. E., and Boehm, F., 1955, *Phys. Rev.*, **97,** 103.

Marty-Wollman, N., 1951, *Ann. Phys., Lpz.*, **6,** 662.

McMuller, C. C., and Johns, M. W., 1953, *Phys. Rev.*, **91,** 418.

Metzger, F. R., and Hill, R. D., 1951, *Phys. Rev.*, **82,** 646.

Miller, L. C., and Curtiss, L. F., 1946, *Phys. Rev.*, **70,** 983.

Muller, D. E., Hoyt, H. C., Klein, D. J., and DuMond, J. W. M., 1952, *Phys. Rev.*, **88,** 775.

Richmond, R., Grant, P. J., and Rose, H., 1952, *Proc. Phys. Soc.* A, **65,** 484.

Rutherford, E., and Andrade, E. N. da C., 1914, *Phil. Mag.*, **27,** 854, **28,** 263.

Siegbahn, K., 1947, *Ark. Mat. Astr. Fys.*, **34 b,** No. 7.

Steffen, R. M., 1951, *Phys. Rev.*, **82,** 824.

Wu, C. S., Havens, W. W., and Rainwater, L. J., 1948, *Phys. Rev.*, **74,** 1248.

The Structure of Cold Worked Gold
I: A Study by Electron Diffraction

By P. B. HIRSCH*, A. KELLY*‡ and J. W. MENTER†‡

* Crystallographic Laboratory, Cavendish Laboratory, Cambridge

† Research Laboratory for the Physics and Chemistry of Surfaces, Department of Physical Chemistry, Cambridge

Communicated by W. H. Taylor; MS. received 11*th February* 1955, *and in amended form* 7*th July* 1955

Abstract. The shape, number and arrangement of spots are studied on electron diffraction photographs of beaten gold foil. Various disturbing effects, including preferred orientation and refraction, on the deduction of particle size by spot counts are considered. After allowance for these a particle size of 1500 Å ± 50% is obtained.

Streaks and displaced spots are explained in terms of crystals with lamellar shapes and stacking faults on (111) planes. Heavily faulted lamellae of thickness less than 150 Å and lamellar regions of hexagonal close packed gold are both found to exist.

§ 1. Introduction

By using x-ray beams of small diameter it has been possible to obtain 'spotty' diffraction rings from deformed polycrystalline metals, and from such patterns the particle size and distortion could be determined (for a summary of the results see Gay, Hirsch and Kelly 1954). It is difficult to obtain x-ray beams sufficiently narrow to resolve the spots from particles of diameter less than about 5000 Å, and further, for such particles the exposure times become prohibitive. Photographs of beaten gold foil taken with an 8 μ beam show rings which are not quite resolved (figure 1, Plate I), indicating a particle size of about 2000 Å (Kelly 1953). In view of the difficulty of applying the x-ray technique to this range of sizes, the use of electron diffraction has been investigated.

The use of electrons instead of x-rays for diffraction patterns has the advantages that parallel beams of small diameter are readily available and that exposure times are short. A further advantage lies in the possibility of combining electron diffraction with electron transmission microscopy to examine the same area of specimen. Most studies of the fine structure of deformed metals have been confined to the examination of the surface structure by replica methods (Heidenreich and Shockley 1948, Brown 1952, Kuhlmann-Wilsdorf and Wilsdorf 1953). Heidenreich (1949, 1951) has used transmission microscopy and diffraction to study the texture of cold-worked aluminium foil, but no quantitative deductions were made from the diffraction patterns.

In the present work an attempt has been made to interpret the diffraction patterns in greater detail than hitherto, and to correlate the results with those from the micrographs. The effects which have been observed in beaten gold foil are complex, and it has not yet been possible to account for all of them. This paper deals with information derived from the shapes of the spots on the diffraction patterns, and with the determination of particle size from spot counts. In a subsequent paper we shall deal with the micrographs and their correlation with the diffraction patterns.

---

‡ A. K. is now at the Department of Physical Metallurgy, University of Birmingham; J. W. M. is now at Tube Investments Research Laboratory, Hinxton Hall, Cambridge.

§ 2. EXPERIMENTAL

Commercial beaten gold foil (~1000 Å thick) was etched with 10% HNO_3 and 10% KCN or KCN alone until it appeared greyish blue under the microscope and contained many pinholes. Two samples were used, the first 99·9% Au and the second 96% Au, 2% Cu, 2% Ag. The foils were mounted on copper grids and transmission electron diffraction patterns and micrographs were taken from the same areas (diameter 10 μ), using the Metropolitan Vickers EM3 electron microscope operating at 70 kv ($\lambda = 0{\cdot}045$ Å). Figure 2 (Plate I) shows typical electron diffraction patterns.

§ 3. ESTIMATION OF LOWER LIMIT OF PARTICLE SIZE FROM SHAPE OF SPOTS

Assuming extinction and refraction effects to be absent, the shapes of the spots can be interpreted in terms of the spread of the interference function around the reciprocal lattice points as in the case for X-ray diffraction. For electrons, the wavelength spread due to variations in the accelerating voltage (<1 part in 20000 in the EM3 instrument) and in the thermal energies of emission of the electrons can be neglected.

It can be shown (Hirsch 1952) that the tangential width s_T of the spots is given by

$$s_T = \frac{R_0}{\cos 2\theta}\,(d\theta + 2d\phi_3 \sin\theta) \qquad \ldots\ldots(1)$$

where R_0 = specimen–film distance, θ = Bragg angle, $d\theta$ = divergence of incident beam, $d\phi_3 = \lambda dr_3/2\sin\theta$, λ = wavelength and dr_3 = extent of interference function at right angles to plane containing the incident and diffracted rays (figure 3).

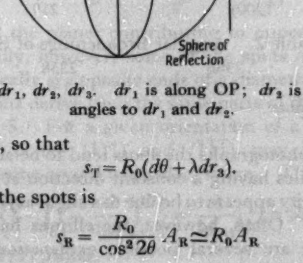


Figure 3. Definition of dr_1, dr_2, dr_3. dr_1 is along OP; dr_2 is along CP; dr_3 is at right angles to dr_1 and dr_2.

For electrons θ is small, so that

$$s_T = R_0(d\theta + \lambda dr_3). \qquad \ldots\ldots(2)$$

The radial width s_R of the spots is

$$s_R = \frac{R_0}{\cos^2 2\theta}\,A_R \simeq R_0 A_R \qquad \ldots\ldots(3)$$

where A_R is a function of $d\theta$, $d\phi_1$ $(=\lambda dr_1/2\cos\theta)$ and $d\phi_2(=\lambda dr_2/2\sin\theta)$ and depends on the relative magnitudes of these quantities. (dr_1 is the width of the interference function along the vector joining the origin of reciprocal space to the reciprocal point, and dr_2 is perpendicular to dr_1 and dr_3, see figure 3.)

If dr_1, dr_2, dr_3 are due entirely to the shapes of the crystals, generally dr_2 will be either greater than, or at least of the same order of magnitude as, dr_1, and hence in either case, as θ is small, $d\phi_2 \gg d\phi_1$. With this assumption there remain only two possibilities for A_R (Hirsch 1952):

if $d\theta < d\phi_2 - d\phi_1$, then $A_R = d\theta + 2d\phi_1 = d\theta + \lambda dr_1$
or if $d\theta > d\phi_2 - d\phi_1$ $A_R = d\phi_1 + d\phi_2 \simeq d\phi_2$.

When θ is small, on averaging the lengths and breadths of the spots around the ring, $dr_1 = dr_3$, and hence $\bar{s}_T = R_0(d\theta + \lambda dr_1)$ in all cases, and either $\bar{s}_R = R_0(d\theta + \lambda dr_1)$ or $\bar{s}_R = R_0 d\phi_2$. In the first case the averaged dimensions of the spots should be equal; in the second case $\bar{s}_T$ should be much greater than $\bar{s}_R$ since $d\theta > d\phi_2 - d\phi_1$.

Table 1 shows a series of measurements for gold foil. The results show that $\bar{s}_T \sim \bar{s}_R$, so that it can be concluded that

$$s_T = s_R = R_0(d\theta + \lambda dr_1) \qquad \ldots\ldots(4)$$

and hence that

$$d\theta < d\phi_2 - d\phi_1. \qquad \ldots\ldots(5)$$

Average values of $d\theta + \lambda dr_1$ for those photographs for which the values of $d\theta$ could be estimated from the manufacturer's data appropriate to the condenser lens current used in the experiments are shown in table 2. It is clear that the major part of the spot width is due to the divergence of the beam. Assuming that the additional broadening is entirely due to crystal shape, a lower limit to the linear dimensions a of the crystal in the plane of the specimen can be found. Thus, for one of the photographs from gold foil, $\lambda dr_1 < 0{\cdot}5 \times 10^{-4}$; hence $\lambda/a < 0{\cdot}5 \times 10^{-4}$ and $a > 1000$ Å. Thus from the spot shapes alone a lower limit of particle size can be estimated.

Table 1. Average Tangential and Radial Spot Lengths

Photograph	Ring	$\bar{s}_T$ (μ)	$\bar{s}_R$ (μ)
2366	(111)	131	134
1070	(200)	170	159
1070	(111)	174	134
1314	(200)	160	153
1315	(200)	211	198
1308	(200)	219	192

Table 2. Angular Broadenings of Spots

Photograph	2366	2358	2364
$d\theta + \lambda dr_1$ (radian × 10^{-4})	2·2	1·6	1·5
$d\theta$ from makers' specifications (radian × 10^{-4})	1	1	1

On almost all the photographs the spots tend to be slightly elliptical, the axes of the ellipses sometimes having a constant direction at all points on the photographs, and the ellipticity appears to be due to a slight asymmetry of the divergence (figure 4 (*a*), Plate I). Often, however, the ellipses have different orientations (figure 4 (*b*)). There are several possible explanations for this observation.

It is likely that dr_2 is generally much greater than dr_1 or dr_3, so that the interference function is in the form of a spike normal to the plane of the foil. If the foil is tilted or bent, the spike may make large angles with the direction of the incident beam, and it is easily seen that this will cause ellipticity of the spots, the direction of the ellipse depending on the angle between the surface of the reflecting sphere and the spike at the point of intersection. Other possibilities are that the particles are in the form of lamellae which make large angles with the plane of the foil, or that stacking faults occur. Further evidence for the existence of thin lamellae and stacking faults is given below.

§ 4. STRAINS IN THE FOIL

By an argument similar to that used to determine the lower limit of particle size, the elastic strains in the plane of the foil cannot exceed a value determined by $d\phi_1$($<0{\cdot}25\times10^{-4}$ radian). Now the range of strains

$$\delta d/d=d\,.\,\delta(1/d)=d\,.\,dr_1=(2d/\lambda)\cos\theta\, d\phi_1,$$

so that for the 200 planes, $\delta d/d<2\times10^{-3}$. This value is of the same order of magnitude as that found for cold-worked metals (Gay, Hirsch and Kelly 1954). It is clear that if such strains were to occur, the broadening produced would be smaller than the divergence of the beam, so that it is not possible to measure the elastic strains. However, if the divergence could be reduced to approximately 10^{-5} radian, the strain broadening could probably be determined.

It is likely that the foil may be elastically bent about an axis in its own plane. If the foil is bent about an axis parallel to dr_1 the spot will be broadened tangentially; the broadening will be $\frac{1}{2}\lambda\alpha/d^2$ where α is the angle of bending, so that $\frac{1}{2}\lambda\alpha/d^2<dr_3=(2\cos\theta/\lambda)d\phi_1$. Hence $\alpha<(2d/\lambda)^2 d\phi_1=4\times10^{-2}$ radian for a high order plane ($d\sim1$ Å). This estimate will be required later.

§ 5. ESTIMATION OF LOWER LIMIT OF FOIL THICKNESS FROM OBSERVATIONS ON PAIRED SPOTS ON OPPOSITE SIDES OF THE RINGS

Spots nearly always occur on opposite sides of the same rings, but the arrangement of spots is generally different on the two sides (e.g. figure 2 (*b*)), and the occurrence of such spots is to be expected from the large angular spread of the particles about their mean direction (see table 3), which is much larger than the angles between the planes contributing to opposite sides of the same ring, i.e. 2θ. Occasionally, however, some of the spots on the (111) rings are reproduced nearly identically at opposite ends of a diameter (figure 2 (*c*)). This suggests that the reflections derive from the same parts of the foil. The explanation follows from figure 5. For a given orientation of a particle the lines dr_2 through the reciprocal points (111) and ($\bar{1}\bar{1}\bar{1}$) may intersect the sphere simultaneously, and the same crystal will reflect twice. This can occur provided

Table 3. Range of Misorientations

Photograph	2358	2366	1070	1307	1314
Angular spread of arcs (°)	30	30	25	20	20

$$\tfrac{1}{2}dr_2>\frac{1}{\lambda}\left\{1-\left(1-\frac{\lambda^2}{{d_{111}}^2}\right)^{1/2}\right\}=\frac{\lambda}{2{d_{111}}^2},\ \text{i.e. } dr_2>\frac{\lambda}{{d_{111}}^2}$$

where d_{111} = spacing of (111) planes.

On the other hand, duplicated spots have not been observed on higher order rings and this fact may be used to deduce an upper limit for dr_2. Paired spots can only be recognized if a whole group of spots appears identically on opposite ends of a diameter of a ring. Apparent absence of paired spots therefore implies that the probability of their occurrence is so small that they cannot be detected. The number of reflections on a ring is proportional to $d\,.\,dr_2 = d\phi_2$. It is easily shown that the number of reflections which will be paired is proportional to $d(dr_2 - \lambda/d^2)$, so that the ratio of the number of paired spots to the total number of spots equals $1 - \lambda/d^2 dr_2$. The conservative criterion may be adopted that duplicated spots are not observed if this ratio is less than about $\frac{1}{3}$, i.e. if $dr_2 < 1\cdot5\lambda/d^2$, where d here is the spacing of that ring showing no duplicated spots which occurs at the smallest angle. Thus, for the specimen discussed above, $dr_2 < 1\cdot5\lambda/d_{111}^2$; it follows that $8\cdot3 \times 10^{-3} < dr_2 < 1\cdot7 \times 10^{-2}$. Hence $dr_2 = 1\cdot25 \times 10^{-2} \pm 35\%$. If dr_2 is entirely due to particle size effects, the thickness t of the crystal is about 85 Å. For most of the gold photographs no duplicated spots are observed on any of the rings, so that $dr_2 < 1\cdot5\lambda/d_{111}^2 = 1\cdot25 \times 10^{-2}$, and the thickness t must exceed 85 Å.

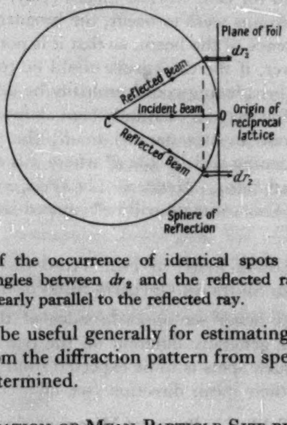


Figure 5. Explanation of the occurrence of identical spots on opposite ends of ring diameters. The angles between dr_2 and the reflected rays have been exaggerated for clarity; dr_2 is nearly parallel to the reflected ray.

These criteria may be useful generally for estimating the thickness or limits of thickness directly from the diffraction pattern from specimens whose thickness cannot be otherwise determined.

§ 6. DETERMINATION OF MEAN PARTICLE SIZE BY SPOT COUNTS

Assuming that the particles are randomly orientated the number of spots N on the ring corresponding to a Bragg angle θ is given by (cf. Hirsch and Kellar 1952)

$$N = \tfrac{1}{2} N_0 p \cos\theta (d\theta + \Delta) \qquad \ldots\ldots (6)$$

where N_0 = number of particles irradiated, p = multiplicity factor and $\Delta = d\phi_1 + d\phi_2$ = angle over which particles reflect. It follows from § 3 that $d\phi_2 \gg d\theta > d\phi_1$; this will be generally true for electron diffraction. Hence, remembering that θ is small,

$$N = \tfrac{1}{2} N_0 p d\phi_2. \qquad \ldots\ldots (7)$$

In the case of x-ray diffraction equation (6) can be used directly to give a fairly accurate value of N_0; however, when equation (7) is applied to electron diffraction patterns, a number of difficulties are encountered. In the first place, a reflection may be split into secondary spots due to refraction of the electron beam in passing through differently orientated facets of the diffracting particle. Secondly, there is an uncertainty in the estimate of $d\phi_2$, which depends on the thickness and distortion of the crystals. Lastly, in specimens used for electron diffraction there is often a pronounced preferred orientation which invalidates the application of equation (7). In order to assess the importance of these effects the validity of equation (7) was tested by observations on diffraction patterns from specimens where the particles responsible for the diffraction spots could be counted on the corresponding micrographs.

§ 7. REFRACTION

Figure 6 (Plate II) shows the diffraction pattern and micrograph of the same area of a sample of MgO on a carbon support. The MgO crystals are seen to be cubes with sides varying between 300 Å and 2000 Å. The diffraction rings consist of clusters of spots of characteristic shape. These groups of spots are due to the refraction of the electrons at the cube surfaces, and as different rays may pass through a different pair of faces, several spots will occur whose grouping depends on the plane of reflection and the orientation of the cubes (Sturkey and Frevel 1945, Cowley and Rees 1947, Honjo 1953). Each group of spots therefore arises from one crystal and should be counted as one spot. In MgO it is easy to recognize these clusters on account of the simple constant symmetrical habit of the crystals, but in a substance for which the particle shapes are irregular, these clusters may not be readily distinguishable and the number of spots observed may be several times greater than the number expected from the simple theory. Even in such a case, however, it is generally possible to estimate the possible error arising from this effect (see § 10). It should be noted that the probability of reflection is not increased by refraction; the probability still depends mainly on $d\phi_2$, and if a crystal reflects at all then it is likely to produce most of the spots caused by refraction. This is confirmed by the fact that most of the clusters on figure 6 consist of many spots.

§ 8. ESTIMATES OF $d\phi_2$

Where possible $d\phi_2$ should be measured experimentally by determining the angular range of reflection of the particles. If this cannot be done, upper and lower limits of $d\phi_2$ may be determined as follows. If the crystals are so thin that extinction is negligible, and if the crystals are perfect, $d\phi_2 = d \, . \, dr_2 = d/t$. With increasing t, $d\phi_2$ decreases, but if extinction becomes important the angle over which the crystal can effectively reflect reaches a lower limit with increasing thickness. This angle can be estimated approximately from the width of the intensity curve for a symmetrical Laue reflection, for which the intensity I is given (Zachariasen 1945) by

$$I = \frac{\sin^2\{A(1+y^2)^{1/2}\}}{1+y^2} \qquad \ldots\ldots(8)$$

where, for electrons,

$$A = \frac{\pi}{\lambda} t \frac{|V_{hkl}|}{E} \qquad \ldots\ldots(9)$$

and

$$y = \frac{E}{V_{hkl}} (\delta\theta) \sin 2\theta \qquad \ldots\ldots(10)$$

(MacGillavry 1940, Heidenreich 1949). V_{hkl} is the *hkl* coefficient of the Fourier expansion of the potential of the crystal, E is the energy of the electron, and $\delta\theta$ is the deviation from the mean Bragg angle θ. Equation (8) shows that the diffraction pattern consists of a series of fringes whose peak intensity, except for the central fringe, is governed by the envelope factor $(1+y^2)^{-1}$. The intensity of this envelope is reduced to half its maximum value at $y=\pm 1$. For thick specimens, and strong reflections, A may be large and several fringes occur within the range $-1<y<+1$; these will be sufficiently strong to give observable reflections on the diffraction pattern. Although the width of the individual fringes may be quite small, the total effective angular range of reflection for a crystal will be of the order of the sum of the widths of the strong fringes. The total width of these fringes will be approximately equal to half the range $-1<y<+1$, i.e. to $\delta y \sim 1$. Hence, from equation (10), the effective angular range of reflection of the thickest crystals will be

$$d\phi_2 \sim \frac{V_{hkl}}{E\sin 2\theta} = \frac{V_{hkl}d_{hkl}}{E\lambda} = \frac{d_{hkl}}{t_0}$$

where $t_0=\lambda E/V_{hkl}$ = 'extinction distance' (cf. Heidenreich 1949, 1951). This value gives the lower limit of $d\phi_2$; if the crystal thickness is less than t_0, or if the crystal is imperfect, $d\phi_2$ will be larger than this limit. V_{hkl} is given by

$$V_{hkl}=\frac{300\boldsymbol{e}d^2_{hkl}}{\pi V}\sum_j (Z_j-f_j)\exp\{2\pi i(hu_j+kv_j+lw_j)\}$$

in volts, where $\boldsymbol{e}$ = electronic charge, V = volume of unit cell, Z_j = atomic number and f_j = scattering factor of jth atom at u_j, v_j, w_j. Table 4 shows the values of V_{hkl} and t_0 calculated for various rings of MgO, and the corresponding upper limit of N_0. It appears that $N_0<2200$.

Table 4. Data for the Determination of N_0 for MgO

(*a*) Upper Limit of N_0

hkl	d_{hkl} (Å)	V_{hkl} (v)	$t_0=\dfrac{\lambda E}{V_{hkl}}$ (Å)	$d\phi_2$ (rad)	p	N	N_0
(111)	2·425	1·02	3090	$7{\cdot}85\times10^{-4}$	8	39	12400
(200)	2·10	8·2	384	$5{\cdot}46\times10^{-3}$	6	36	2200
(220)	1·485	5·78	545	$2{\cdot}72\times10^{-3}$	12	64	3920
(222)	1·212	4·4	715	$1{\cdot}69\times10^{-3}$	8	23	3400

(*b*) Lower Limit of N_0

hkl	Upper limit of $d\phi_2$ (rad $\times 10^{-2}$)	N_0
(111)	2·79	350
(200)	2·79	430
(220)	2·79	382
(222)	2·79	206

(*c*) Mean Value of N_0

$$430<N_0<2200$$
$$N_0=1300\pm 900$$

A different argument must be used to obtain the upper limit of $d\phi_2$. Many of the cubes occur in orientations such that the beam traverses thicknesses from zero upwards, which would suggest the possibility of very large values of $d\phi_2$. However, the intensities would be very low from those parts of the crystal which are very thin. An effective upper limit of $d\phi_2$ is however available from the argument of § 5. Since no spots are paired on opposite sides of the ring,

$dr_2 < 1{\cdot}5\lambda/d_{111}^2$; hence $d\phi_2 < 2{\cdot}79 \times 10^{-2}$ radian. The minimum values of N_0 calculated in this way are shown in table 4. It appears then that the actual value of N_0 should lie between 430 and 2200, so that $N_0 = 1300 \pm 900$. This agrees satisfactorily with the actual number of approximately 1000 crystals on the micrograph. This example illustrates the kind of accuracy which can be expected from this method. Sometimes it is possible to estimate $d\phi_2$ more accurately, as will be indicated below.

§ 9. PREFERRED ORIENTATION

A further difficulty in the accurate estimation of N_0 arises if there is preferred orientation in the specimen. This can often be very marked in gold (figures 2 (*b*), (*c*)). The spots are clustered into arcs and the range of misorientation obtained from the angular extent of the arcs is shown in table 3. The results are of the same order of magnitude as those found using the x-ray microbeam technique for most deformed metals (Gay and Kelly 1953 a, b). The patterns suggest that the areas irradiated consist mainly of one original crystal.

Many of the patterns show a preferred orientation of the (100) plane parallel to the plane of the foil (figure 2 (*b*)). Occasionally other orientations are observed, for example that of figure 2 (*c*). It is likely that these other orientations arise from a bending or tilting of the foil so that its plane is not normal to the direction of the incident beam.

The likely error in N_0 arising from preferred orientation is illustrated by observations on a thin layer of tin evaporated on to a collodion support. Both the diffraction pattern and micrograph (figure 7, Plate II) suggest that the refraction effect is small. However, the numbers of spots on the different rings did not agree with the trend of the multiplicity factor. The variation of the number of spots suggested a preferred orientation of the crystals with [001] parallel to the plane of the specimen. This was confirmed by tilting the specimen; the preferred orientation is however not very pronounced, so that at least an approximate value of N_0 may be determined by averaging the different values obtained. In this case the extinction distances calculated for the different rings all exceeded the thickness of the specimen which was found by interferometry to be 200 ± 20 Å. This distance therefore determines the lower limits of $d\phi_2$; with these values the mean upper limit of N_0 is found to be 7400 (see table 5). No paired spots are observed so that the upper limit of $d\phi_2$ is $2{\cdot}29 \times 10^{-2}$ radian. The mean lower limit of N_0 is found to be 4100 (see table 5). The average of these two limits is 5750; the number of crystals apparent on the micrograph is about 3000, although, if there are smaller crystals between the larger ones, the true figure may be considerably greater. In view of the uncertainty of this figure and the error due to the preferred orientation, this agreement to within a factor of 2 is satisfactory.

§ 10. PARTICLE SIZE OF GOLD

While many of the photographs from gold foil showed the cube texture (figure 2 (*b*)) in certain cases the crystals seemed to occur in reasonably random orientations (figure 2 (*a*)). Table 6 shows the data obtained from two photographs. No significant differences were observed between the numbers of spots obtained from the two types of gold of different purity examined. Duplicated spots were not detected, and as before, using this criterion and the extinction distance, upper and lower limits of N_0 were calculated (table 6). In order to reduce the error due to preferred orientation the resultant values were averaged to give a value of $N_0 = 4700 \pm 2300$. In this case, however, there remains the possibility of bending of the foil and this can cause a considerable increase in the effective value of $d\phi_2$. Using the upper limit to the angle of bending α found in § 4, $N_0 \sim 4000 \pm 3000$; thus the possibility of bending can reduce the accuracy of the estimate considerably. In order to overcome this difficulty, the angular range of reflection was determined directly by tilting the electron gun in the microscope through a range of angles, and observing the changes in the diffraction pattern. In this way it was found that $d\phi_2 \sim 10^{-2}$, so that the mean value of N_0 from both photographs is of the order of 6900.

Table 5. Data for the Determination of N_0 for Sn

$t = 200 \pm 20$ Å

(*a*) Upper Limit of N_0

hkl	d_{hkl} (Å)	$d\phi_2$ (rad × 10^{-2})	*p*	*N*	N_0	
(200)	2·95	1·475	4	358	12100	Mean N_0 7400
(101)	2·81	1·40	8	249	4450	
(220)	2·08	1·04	4	211	10150	
(211)	2·02	1·01	16	232	2870	

(*b*) Lower Limit of N_0

hkl	Upper limit of $d\phi_2$ (rad × 10^{-2})	N_0	
(200)	2·29	7820	Mean N_0 4100
(101)	2·29	2720	
(220)	2·29	4600	
(211)	2·29	1270	

(*c*) Mean Value of $N_0 = 5750$
(the possible error of this value due to preferred orientation is uncertain)

Table 6. Data for the Determination of Particle Size in Au

(*a*) Upper Limit of N_0

					Photo 3064				Photo 2364		
hkl	d_{hkl} (Å)	V_{hkl} (v)	$t_0 = \frac{\lambda E}{V_{hkl}}$ (Å)	$d\phi_2$ (rad)	*p*	*N*	N_0		*N*	N_0	
(111)	2·35	22·5	140	$1{\cdot}68 \times 10^{-2}$	8	214	3180	Mean N_0 6820	333	4950	Mean N_0 7310
(200)	2·035	20·2	156	$1{\cdot}305 \times 10^{-2}$	6	238	6080		257	6560	
(220)	1·44	14·32	220	$6{\cdot}55 \times 10^{-3}$	12	432	11000		410	10420	
(311)	1·228	12·15	257	$4{\cdot}75 \times 10^{-3}$	24	400	7010				

(*b*) Lower Limit of N_0

hkl	Upper limit of $d\phi_2$ (radians)	Photo 3064 N_0		Photo 2364 N_0	
(111)	$2{\cdot}87 \times 10^{-2}$	1860	Mean N_0 2074	2900	Mean N_0 2753
(200)	$2{\cdot}87 \times 10^{-2}$	2765		2980	
(220)	$2{\cdot}87 \times 10^{-2}$	2510		2380	
(311)	$2{\cdot}87 \times 10^{-2}$	1160			

(*c*) Mean Value of N_0

From photograph 3064	$2074 < N_0 < 6820$	or $N_0 = 4500 \mp 2400$
From photograph 2364	$2753 < N_0 < 7310$	or $N_0 = 5000 \pm 2300$

Mean value from both photographs $N_0 = 4700 \pm 2300$
Mean value of N_0 allowing for possible bending 4000 ± 3000
Mean value based on experimental value of $d\phi_2 = 10^{-2}$, 6900
Diameter of aperture 10 μ
Mean diameter of particle in plane of foil $a = 1070$ Å
Allowing for refraction, $a \sim 1500$ Å (± 50%)

In the case of the foil, $N_0 = A/a^2$ where A is the diffracting area of the foil in the beam and a the diameter of the particles in the plane of the foil. For the two photographs discussed above the whole area in the aperture appears to be diffracting and the area of holes is negligible. Hence A is equal to the area (diameter $10\,\mu$) selected by the aperture. Using the mean value of N_0 deduced above, a is found to be 1070 Å.

It is possible that owing to the refraction effect (see above) several spots are produced by one particle. Occasionally ring splitting is observed as in figure 2 (*b*); this effect could be due to refraction, although stacking faults could also produce it. In figure 2 (*a*) no such splitting occurs and the effect of refraction is negligible. In figure 2 (*b*) frequent splitting is observed and the number of spots could be increased by at most a factor of 3 by this effect. This leads to an upper limit of particle size determined from this photograph of 1850 Å. We may take the average particle size to be of the order of 1500 Å ± 50%.

§ 11. STREAKS

It has already been mentioned that diffuse streaks pass through many of the spots (figures 2 (*a*), 4 (*b*), (*c*)). These streaks occur both with etched and unetched specimens of pure and impure gold. The nature of the streaks suggests that they are due to rods in reciprocal space. In general the reflecting sphere intersects the rods at an arbitrary angle, and only if the orientation of the crystal is such that the rods are tangential to the reflecting sphere can the direction of the rods be determined. It is reasonable to assume that this is the case for the longest streaks observed. The angle between the streak and the normal to the reflecting plane then gives directly the orientation of the rod. Table 7 shows experimentally observed angles for the (111) and (200) rings, and the values expected if the rods are along [111]. The two sets of values are in good agreement, showing that the rods are along [111]. Very occasionally other streaks are observed, e.g. radial to (111) and (200). The rods in reciprocal space may be due either to crystals in the form of thin plates (i.e. lamellae, generally parallel to the (111) plane) or to stacking faults on (111) (Edwards and Lipson 1942, Wilson 1942, 1949, Barrett 1950, Paterson 1952, Warren and Warekois 1953). Experimentally it is found that streaks radial to the (111) ring occur only very rarely compared with the other orientations given in table 7. This suggests strongly that many of the streaks are due to stacking faults, for stacking faulting on a given set of (111) planes cannot affect the reflection from that set. Further confirmation that many of the streaks are due to stacking faults lies in the fact that they are usually associated with displaced spots.

Table 7. Angles between Streaks and Normals to Reflecting Planes

hkl	Observed angles	Angle between *hkl* and (111)
(111)	70°, 70°, 70°, 65°, 70°	70° 32′
(200)	50°, 50°, 60°, 50°, 50°	54° 44′

It is convenient to discuss first the occasional radial streaks or radially elongated spots on (111) which are probably due to lamellae. Some of the streaks appear to consist of regularly spaced faint spots. These faint spots are likely to be due to the subsidiary maxima of the interference function (Rees and Spink 1950, Uyeda, Ichinokawa and Fukano 1954). Nearly all these spotty streaks are unsymmetrical about the main spot. Such unsymmetrical streaks can be explained if the rods in reciprocal space are not tangential to the reflecting sphere, and if at the same time the crystal is bent (Uyeda 1953). It is not possible to obtain an accurate estimate of the lamellar thickness from the spacings of the spots, as the exact orientation of the rods is unknown, but an approximate upper limit to the thickness can be deduced and is found to be of the order of 300 Å. The direction of the streaks suggests that the lamellae are generally parallel to (111).

Next we consider the streaks due to stacking faults; these can be interpreted in terms of the theory of Paterson (1952) who calculated the intensity distributions from crystals with a random distribution of deformation or growth faults. A single deformation fault causes the stacking sequence of (111) planes to change from ABCABCABC... to ABABCABC... whereas a single growth fault produces the sequence ABCABACBACBA..., which corresponds to the stacking sequence across a face-centred cubic twin-boundary.

In both types of faulting the reflections are broadened, streaks are produced and the position of the peak is displaced, the broadening and displacement depending on the probability of occurrence of faults. Most of the streaks found on the photographs do not decrease rapidly in intensity with distance from the peak, as would be expected on Paterson's theory (figure 4 (*b*), (*c*)). On the contrary, one, or sometimes several irregularly spaced, weak diffuse maxima appear along these streaks. This means that the faulting giving rise to these streaks is not random but that in local regions in the crystal the probability of faulting is very high, thus producing a diffuse displaced spot. The appearance of several faint maxima suggests the existence of several regions with different probabilities of faulting. A further proof that most of the long streaks observed are of this nature is the fact that the streaks are not symmetrical about the spots; for example, for the (111) ring the streaks run generally towards (200) and vice versa (see figure 2 (*a*)). If the faulting were random the streaks should be symmetrical about (111) and (200), while the displacements of the peaks from the faulted regions should be from (111) to (200) and vice versa (Paterson 1952). The fact that the streaks associated with the (111) and (200) rings run predominantly towards each other is also further strong evidence that many of these streaks are due to stacking faults, and not due to thin lamellae.

It is possible to deduce the faulting probability of the heavily faulted regions from the displacement of the diffuse peaks. If the probability of faulting $\alpha = \frac{1}{2}$ the peak occurs at the position of the reflection from a hexagonal close-packed crystal with l odd; further increase of α results in the movement of the reflection to the position expected from the face-centred cubic crystal in the twin orientation. It appears that many of the streaks and displaced spots correspond to values of α between $\frac{1}{4}$ and $\frac{1}{2}$, while a few streaks suggest values of $\alpha > \frac{1}{2}$; in almost every case where a streak starts on (111), a reflection occurs on (200) in the correct orientation for twinning, but no case has been clearly identified where the streak is continuous from a spot on (111) to another spot on (200). The evidence on twinning remains therefore inconclusive; in particular, no uniquely identified twins of appreciable thickness have been observed. All the available evidence thus points to the conclusion that the faulted regions are a consequence of deformation, and are not due to growth faults originally present in the crystals; for such faults would naturally lead to continuous streaks and spots symmetrically placed about the hexagonal close-packed reflection position, for example the $(10\bar{1}1)$ reflection between the (111) and (200) rings. No such pairs of spots occur. The integrated intensity of the diffuse streaks is generally less than about 1/10 of that of the strong undisplaced spots; the breadth of the streak (measured normal to its length) is of the same order as that of the spots. This indicates that the faulting occurs in lamellae parallel to (111) extending over the whole width of the crystal, i.e. $\sim$1500 Å, but that the thickness of the lamellae is generally less than 150 Å.

§ 12. Extra Spots

It has already been mentioned that spots occur which do not lie on the rings of the face-centred cubic lattice (figure 2 (*a*)). Some of these are displaced by random amounts and can be accounted for by stacking faults. In addition extra spots occur on well-defined rings; these rings occur at the positions expected from a hexagonal close-packed structure with axial ratio 1·633 (see table 8), i.e. a structure in which the close-packed layers are arranged in the sequence ABAB... rather than ABCABC....[†]

Table 8. Diffraction Rings due to Hexagonal Close-packed Lattice

hkil	*d*(obs) (Å)	*d*(calc) (Å)
$(10\bar{1}0)$	2·50	2·525
$(10\bar{1}2)$	1·71	1·72
$(10\bar{1}3)$	1·33	1·33
$(20\bar{2}2)$	1·10	1·105
$(20\bar{2}3)$	0·97	0·978

The diffraction rings from a hexagonal close-packed lattice fall into three groups: (1) $h+2k=3N+1$, l odd (where N is any integer), (2) $h+2k=3N+1$, l even, (3) $h+2k=3N$, l even. If a face-centred cubic crystal suffers random deformation faulting, and if $\alpha=\frac{1}{2}$, the lines of the face-centred cubic lattice move to the positions of the lines of group (1); the lines of group (2) do not occur. The occurrence of lines $(10\bar{1}2)$, $(20\bar{2}2)$ (see table 8) which belong to this set proves that a true hexagonal phase exists. This phase could also be produced by faulting (but not by random faulting); it corresponds to a stacking fault occurring on every second plane of the face-centred cubic arrangement. It may be added here that Troiano and Tokich (1948) found that the cubic phase of cobalt could be transformed by plastic deformation into the hexagonal phase.

The spots of group (2) are weak and in the form of elongated ellipses. Measurements of the orientations of the longer ellipses indicate elongation in reciprocal space along [111]. (Thus, for example, the angle between the streaks on the hexagonal close-packed lattice line $(10\bar{1}2)$ on figure 4 (*c*), between the (200) and (220) rings, and the line joining the centre of the photograph to the streaks, is approximately 35°, in good agreement with the calculated angle of 39° between $(10\bar{1}2)$ and (0001).) This could be due to deformation faulting of the phase, or due to the lamellar shape of the crystals, or probably to both. The width of the spot normal to its length is of the same order as that of the face-centred cubic lattice spots, indicating that the hexagonal close-packed phase extends over the whole width of the crystals. Rough estimates of the intensities of the stronger spots suggest that the thickness of these hexagonal close-packed regions is about one-third of that of the face-centred cubic crystals (i.e. ~500 Å or less). Comparison of the number of spots on the hexagonal close-packed diffraction rings with those on the face-centred cubic rings with the same multiplicity factor shows that about one crystal in ten is hexagonal close-packed. The appearance of very weak spots on very long exposures shows, however, that many more smaller hexagonal close-packed crystals occur, the intensities from which are too weak to be measured on the photographs.

[†] We are indebted to Dr. J. Sawkill for drawing our attention to these hexagonal close-packed lattice rings.

The spots on rings of group (1) are on the whole considerably more streaky than those of group (2). This difference cannot be accounted for by lamellar shape, growth or deformation faulting of the hexagonal phase; as many of the streaks are not symmetrical the explanation is doubtless that they are produced by random deformation faulting of the face-centred cubic phase with $\alpha=\frac{1}{2}$.

§ 13. CONCLUSIONS

The results show that it is possible to obtain information about the structure of cold-worked gold from a detailed study of the number, shape and arrangement of spots on the transmission electron diffraction pattern from a thin foil. A lower limit of particle size (~1000 Å) is deducible from spot shape alone and an order of magnitude estimate of the average size (~1500 Å) can be made by counting spots on a given diffraction ring. The accuracy of this estimate is limited by effects arising from preferred orientation, refraction, stacking faults and strains in the foil. By using a selected area of diameter $2\,\mu$ instead of $10\,\mu$, it should be possible to obtain patterns with the same number of spots as on figure 2, from particles one-fifth of the size of the gold particles, i.e. from particles of mean linear dimension ~300 Å. The accuracy of particle size and strain determination could be improved considerably with a beam divergence of 10^{-5} radian. Although the estimate of size obtained is very approximate, the alternative method of determining particle size by observations on the breadths of the diffraction lines will be even more inaccurate in cases where any of the disturbing effects mentioned above are present. The method is complementary to electron microscopy, where, as in the case of gold foil, it is sometimes difficult to determine the size of the particles by direct observation.

The particle size of beaten gold foil determined by electron diffraction is in good agreement with that suggested by the x-ray patterns (~2000 Å) (Kelly 1953), and is considerably smaller than that found for other cold-worked metals ($\sim 1\,\mu$) (Gay, Hirsch and Kelly 1954).

The particles have generally a preferred orientation with a [100] axis normal to the foil; the spread of orientation (~30°) in an original grain is of the same order as found for other metals (e.g. Gay, Hirsch and Kelly 1954). A study of identical spots on opposite ends of diameters of rings can give information about the thickness of the specimen.

Streaks and displaced spots have been interpreted in terms of stacking faults and (less frequently) crystals in the form of thin lamellae. Heavily faulted ($\alpha\sim\frac{1}{2}$) lamellae occur parallel to (111), and their thickness is less than about 150 Å. A true hexagonal phase in the form of lamellae parallel to (111), of thickness about 500 Å or less, is also found. These lamellae extend over the whole width of the face-centred cubic crystal. No definite evidence of twins of appreciable thickness has been obtained.

These deductions from the electron diffraction patterns can be utilized fully in understanding the nature of plastic deformation in gold only when taken in conjunction with the results of the electron microscopic studies, to be described in a subsequent paper.

ACKNOWLEDGMENTS

Our thanks are due to Professor Sir Lawrence Bragg, Dr. W. H. Taylor and Dr. F. P. Bowden, for their constant advice and encouragement. We are also grateful to Professor N. F. Mott and Dr. J. Sawkill for some helpful discussions and to Mr. C. Whiley of Messrs Geo. Whiley & Co. Ltd., for supplying us with samples of gold leaf. Two of us (P. B. H. and J. W. M.) thank the Managers of the I.C.I. Fellowships Fund of the University of Cambridge for the award of Fellowships, and one of us (A. K.) acknowledges the award of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

BARRETT, C. S., 1950, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **188**, 123.
BROWN, A. F., 1952, *Advanc. Phys.*, **1**, 427.
COWLEY, J. M., and REES, A. L. G., 1947, *Proc. Phys. Soc.*, **59**, 287.
EDWARDS, O. S., and LIPSON, H., 1942, *Proc. Roy. Soc.* A, **180**, 268.
GAY, P., HIRSCH, P. B., and KELLY, A., 1954, *Acta Cryst., Camb.*, **7**, 41.
GAY, P., and KELLY, A., 1953 a, *Acta Cryst., Camb.*, **6**, 165; 1953 b, *Ibid.*, **6**, 172.
HEIDENREICH, R. D., 1949, *J. Appl. Phys.*, **20**, 993; 1951, *Bell Syst. Tech. J.*, **30**, 867.
HEIDENREICH, R. D., and SHOCKLEY, W., 1948, *Report of a Conference on Strength of Solids* (London : Physical Society).
HIRSCH, P. B., 1952, *Acta Cryst., Camb.*, **5**, 168.
HIRSCH, P. B., and KELLAR, J. N., 1952, *Acta Cryst., Camb.*, **5**, 162.
HONJO, G., 1953, *J. Phys. Soc. Japan*, **8**, 776.
KELLY, A., 1953, *Ph.D. Thesis*, University of Cambridge.
KUHLMANN-WILSDORF, D., and WILSDORF, H., 1953, *Acta Met.*, **1**, 394.
MACGILLAVRY, C. H., 1940, *Physica*, **7**, 334.
PATERSON, M. S., 1952, *J. Appl. Phys.*, **23**, 805.
REES, A. L. G., and SPINK, J. A., 1950, *Acta Cryst., Camb.*, **3**, 316.
STURKEY, L., and FREVEL, L. K., 1945, *Phys. Rev.*, **68**, 56.
TROIANO, A. R., and TOKICH, J. L., 1948, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **175**, 728.
UYEDA, R., 1953, *Proc. Imp. Acad. Japan*, **29**, 331.
UYEDA, R., ICHINOKAWA, T., and FUKANO, Y., 1954, *Acta Cryst., Camb.*, **7**, 216.
WARREN, B. E., and WAREKOIS, E. P., 1953, *J. Appl. Phys.*, **24**, 951.
WILSON, A. J. C., 1942, *Proc. Roy. Soc.* A, **180**, 277; 1949, *X-Ray Optics* (London : Methuen).
ZACHARIASEN, W. H., 1945, *The Theory of X-ray Diffraction in Crystals* (New York : Wiley).

RESEARCH NOTES

D.C. and Pulsed Emission from L-Cathodes

By I. BRODIE
Research Laboratories of The General Electric Company, Limited, Wembley, Middx.

MS. received 11*th July* 1955

Wright (1949 a, b) and Feaster (1949) have shown that the difference between d.c. and pulsed emission in an oxide cathode is due to a decay effect with a time constant of the order 10^{-3} to 10^{-2} second at temperatures of about 1000°c and longer at lower temperatures, the pulsed emission being about two orders of magnitude greater than the d.c. emission. They were further able to show that if perfectly clean anode conditions are obtained the d.c. and pulsed emissions are the same, thus indicating that the decay effect is due mainly to ‘ poisoning ’ of the emission by ions returning from the anode or anode–cathode space. Bombardment of the cathode by positive ions denudes the surface of the oxide cathode of its monolayer of barium (Wright 1953). An alternative explanation for the difference between pulsed and d.c. emission from an oxide cathode given by Nergaard (1952) is that it is due to diffusion of barium ions away from the surface whilst emission is being drawn.

The observation reported here is that a difference between pulsed and d.c. emission exists also in L-type cathodes (Lemmens, Jansen and Loosjes 1950) though to a smaller degree than in oxide cathodes. Indications of the effect existing in L-cathodes have been observed by Schaefer and White (1952) who noted that if the temperature of an L-cathode is lowered, the emission at a constant saturating voltage falls to a minimum and then gradually rises to a steady value. This is presumably the cathode recovering from the greater ‘ poisoning ’ effects experienced at higher current densities.

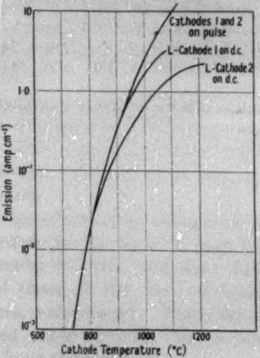


Figure 1.

Figure 1 shows plots of d.c. and pulsed emission against temperature for an L-cathode made by Philips (Eindhoven) and one made by the author. Each of the test diodes had a 7 mm diameter cathode and a 2·0 cm diameter copper (O.F.H.C.) anode placed 1 mm from the cathode. Direct water cooling of the copper anode enabled up to 3 A cm^{-2} d.c. to be drawn from the cathodes. This limit was set by the fact that when the anode dissipation was greater than a kilowatt local boiling of the cooling water occurred however great the flow. Since the thickness of the anode was 7 mm this implies that during the d.c. experiments the temperature of the surface of the anode was never greater than 160°C. The valves were baked at 400°C for half an hour, the cathodes outgassed and a barium getter evaporated on to the bulb before sealing off. Temperature measurements were made with an optical pyrometer focused on small holes drilled into the sides of the porous tungsten discs and were taken, after correction for absorption in the glass envelope, as true black-body temperatures. After activation of the cathode, the valves were aged by running for many hours with 500 volts p.d. between anode and cathode and an emission near the value of the space charge limited current until repeatable results were obtained. The main difference between the two cathodes was in the porosity and thickness of the tungsten barrier. Defining porosity by $P=(1-V_T)\times 100$ where V_T = volume of tungsten per unit volume of material, the values of P for the cathodes used were: cathode 1, $P=46\%$, thickness 1 mm; cathode 2, (Philips) $P\sim 30\%$, mean thickness 0·7 mm. The resistance of similar porous tungsten plugs to air flow was also measured and found to be approximately in the ratio 1 : 1·5. The method of measuring the d.c. emission from the cathodes was to plot the current against the voltage on log–log graph paper and to take the current at the break away from the straight line corresponding to the space charge limited condition. At high emission densities (greater than 0·5 A cm^{-2}) the results are slightly modified by work function cooling of the cathode, but the break away point is still clearly defined and the temperature of the cathode was measured when drawing that current.

It is seen from figure 1 that the pulse emission values for the two types of cathode are the same but the d.c. emission curves depart from the pulse curve at about 500 mA cm^{-2} and 100 mA cm^{-2} respectively.

Lemmens *et al.* (1950) used a molybdenum anode inside the valve envelope which could be cleaned thoroughly by eddy current heating and, up to the limit set by back heating of the anode, detected no difference between pulsed and d.c. emission. Espersen (1952) also used a water cooled copper anode but made no mention of this decay effect.

In an L-type cathode the only reasonable explanation for the difference between pulsed and d.c. emissions is 'poisoning' by positive ion or gas bombardment of the cathode surface which removes part of the monolayer of barium on tungsten. In the cathode where the rate of replenishment of barium is greatest the effect of such barium removal is less marked.

A corollary of this work is that in demountable systems or valves where there is likely to be excessive ion bombardment of the cathode, it is better to use the more open porous tungsten barrier. This enables the emission to be maintained at the cost of somewhat greater barium evaporation and shorter ultimate life.

The poisoning effect due to ion bombardment has also been observed by us on the electron image of an impregnated L-type cathode (Levi 1953). This cathode was activated by heating only. When the accelerating and modulating voltages were applied, the cathode image first gave a uniform appearance (figure 2, Plate) and then the centre portion became much darker in the course of a few seconds (figure 3, Plate). The centre could always be re-activated by removing the h.t. supply for a short time.

REFERENCES

ESPERSEN, G. A., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 248.
FEASTER, G. E., 1949, *J. Appl. Phys.*, **20,** 415.
LEMMENS, H. J., JANSEN, M. J., and LOOSJES, R., 1950, *Philips Tech. Rev.*, **11,** 341.
LEVI, R., 1953, *J. Appl. Phys.*, **24,** 223.
NERGAARD, L. S., 1952, *R.C.A. Rev.*, **13,** 464.
SCHAEFER, D. L., and WHITE, J. E., 1952, *J. Appl. Phys.*, **23,** 669.
WRIGHT, D. A., 1949 a, *Proc. Phys. Soc.* B, **62,** 188; 1949 b, *Ibid.*, **62,** 398; 1953, *Proc. Instn Elect. Engrs*, Pt. III, **100,** 125.

The Measurement of the Coherence of a Light Wave by means of a Lloyd Mirror

By A. H. McILRAITH
Physics Department, Canterbury University College, Christchurch, New Zealand

MS. received 3rd *January* 1955 *and in amended form* 20th *July* 1955

A THEORETICAL treatment by Hopkins (1951), confirmed experimentally by Baker (1953) and by Arnulf, Dupuy and Flamant (1953), shows that the coherence of a wave front can be found by performing a Young's interference experiment on it and measuring the fringe visibility for a succession of pin hole (or slit) separations. The purpose of this note is to draw attention to the advantages offered by a Lloyd mirror experiment as an alternative method for measuring coherence.

In the Young experiment, small sections of the wave front at $+\omega$ and $-\omega$ (figure 1 (*a*)) are isolated and allowed to interfere. The visibility V_ω of the resultant fringe system is then equal to $|\gamma_\omega|$ the coherence or correlation in phase between the points $\pm\omega$. If, instead, a Lloyd mirror (figure 1 (*b*)) is employed with $d/2$ small compared with the distance of the plane of observation from the source, the rays $\pm\omega$ are again able to interfere. As in the Young experiment $|\gamma_\omega|$ is found by measuring the fringe visibility at ω in the plane of observation. However, as the intensities of adjacent maxima (and minina) are unequal we put

$$V_\omega' = [2I_2 - (I_1 + I_3)]/[2I_2 + I_1 + I_3]$$

where I_1 and I_3 are the intensities of adjacent minima and I_2 is the intensity of the intervening maximum. According to calculations following the method of Hopkins (1951), V_ω' differs from V_ω by less than 0·001 in practice. In addition, as

$$\arg \gamma_\omega = 2\pi N - (2\pi d \sin \omega)/\lambda$$

where N is the order of interference, $\arg \gamma_\omega$ can readily be obtained by finding the displacement of the fringes with respect to the mirror, or better, by finding their displacement with respect to a set of reference fringes (McIlraith 1955).

If unpolarized light is used V_ω' fails to equal $|\gamma_\omega|$ at other than very small angles owing to the difference between the phase changes on reflection for the components polarized parallel and perpendicular to the plane of incidence. Thus using the optical constants of aluminium (Schulz 1954, Schulz and Tangherlini 1954) we find that when $\omega = 2°$, $6°$ and $30°$, $V_\omega'/V_\omega = 0{\cdot}98$, $0{\cdot}90$ and $0{\cdot}27$ respectively. Therefore, unless ω is small, the light must be polarized. Readily available polarizers should be adequate for this purpose as $V_\omega'/V_\omega \geqslant 0{\cdot}98$ for $\omega \leqslant 30°$ when 1% of the light is unpolarized.

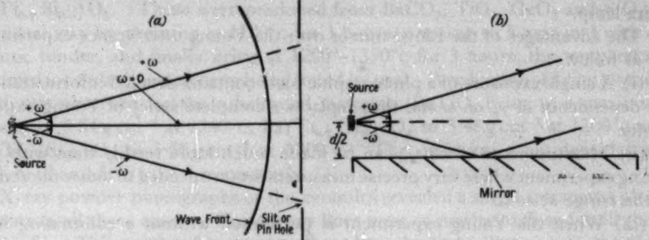


Figure 1. Measurement of wave coherence.

The fringe visibility is almost independent of the coefficient of reflection of the mirror and its surface roughness. Thus when $|\gamma_\omega| = 1$, V_ω' drops to 0·98 when either the coefficient of reflection drops to 0·67 or when 4% of the light is scattered.

When visibility is measured at large values of ω the order of interference may be as high as one thousand thus necessitating the use of spectrally pure light. To find the suitability of a particular source, pieces of glass of different thickness were placed in the path of the rays reflected from the mirror at grazing angles. The fringes formed adjacent to the edge of the shadow of the glass were matched by eye with fringes of the same intensity and spacing in the unobstructed field where the visibility was known. This, with the optical thickness of the glass, made it possible to find the visibility due to spectral impurity alone as a function of the order of interference to within 5%. The results obtained using the mercury green line from the Philips H.P. lamp and the G.E.C. 250 watt Osira lamp with their outer bulbs removed are as follows:

Lamp	Current (amps)	A.C. Voltage (volts)	Order of interference at $V=0{\cdot}5$	$V=0{\cdot}9$	$V=0{\cdot}99$
H.P.	0·80	57	3100	750	130
Osira	2·5	94	12000	2800	500

When the intensity of the H.P. lamp is reduced by a factor of five by cooling the order of interference is increased by a factor of three. Hence the Osira lamp and the cooled H.P. lamp are suitable sources for this work at values of ω as large as 30°.

As it was not possible to measure intensity and hence visibility reliably with the equipment available, an experimental test reported earlier (McIlraith 1955) of the relationship between the width of a virtual slit formed by reflection at a cylindrical surface and the position of the first zero of visibility was repeated with the source close behind the object slit. It was then found that the width of the object slit deduced from the position of the above zero agreed with the directly measured value to within the experimental error of 2%. This also shows that reflecting cylinders are convenient tools for transforming narrow angle diffraction patterns into wide angle patterns. The aberrations thus introduced do not affect the measurement of arg γ_ω provided the reference and unknown sources occupy the same position.

Typical exposure times when a Wratten 77A filter and a polaroid polarizer are used with a virtual slit 1 cm high, 10^{-4} cm wide and 20 cm from a Selochrome plate are ten minutes for the uncooled H.P. lamp and fourteen minutes for the Osira lamp.

The advantages of the Lloyd mirror over the Young interference experiment are as follows:

(i) A single exposure of a photographic plate contains detailed information on the dependence of γ_ω on ω and thus replaces a complete series of Young experiments.

(ii) Measurements of arg γ_ω can be made much more readily than with the Young experiment where very precise measurements are needed to locate the centre of the fringe system.

(iii) When the Young experiment is performed without a collimating lens light losses due to diffraction are such that the Lloyd mirror fringes are much more intense despite the use of a filter and a polarizer. When a collimating lens is used the Lloyd fringes are at least equal in intensity to the Young fringes if the source contains narrow spectral lines.

(iv) The difficulties associated with the manufacture of the two fine parallel slits, which replace the pin holes in the Young experiment when more intense fringes are required, are avoided.

ACKNOWLEDGMENTS

I am indebted to Imperial Chemical Industries, New Zealand Limited, and to the University of New Zealand for the award of Research Fellowships.

REFERENCES

ARNULF, A., DUPUY, O., and FLAMANT, F., 1953, *Rev. Opt.* (*théor. instrum.*), **32,** 10.
BAKER, L. R., 1953, *Proc. Phys. Soc.* B, **66,** 975.
HOPKINS, H. H., 1951, *Proc. Roy. Soc.* A, **208,** 263.
MCILRAITH, A. H., 1955, *Proc. Phys. Soc.* B, **68,** 649.
SCHULZ, L. G., 1954, *J. Opt. Soc. Amer.*, **44,** 357.
SCHULZ, L. G., and TANGHERLINI, F. R., 1954, *J. Opt. Soc. Amer.*, **44,** 362.

Replacement of Ti in $BaTiO_3$ Ceramic by Si and Ge

BY K. W. PLESSNER AND R. WEST
British Dielectric Research Ltd., 38 Wood Lane, London W.12

Communicated by J. B. Birks ; MS. received 28*th July* 1955,
and in final form 16*th August* 1955

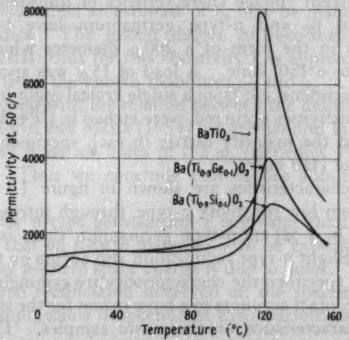
WHEN some fraction of the Ti ions in barium titanate is replaced by *larger* quadrivalent ions, such as Sn and Zr, the main effect is to displace the ferroelectric transition to lower temperatures (Dungen *et al.* 1952

Coffeen 1953, Smolenskiĭ *et al.* 1954). We have replaced some of the Ti ions by the *smaller* quadrivalent ions Si and Ge, and examined the resulting ceramics electrically and structurally. The National Lead Co. (1951) have previously examined the effect of up to 2% SiO_2 added as an impurity to $BaTiO_3$ and found that the permittivity peak was depressed without being displaced. They did not, however, study the effect of replacing substantial fractions of the Ti ions by Si ions.

In the present work three compositions were prepared, using raw materials from the same source and the same processing technique. The ceramic compositions tested had the molecular formulae: $BaTiO_3$, $Ba(Ti_{0\cdot9}Ge_{0\cdot1})O_3$ and $Ba(Ti_{0\cdot9}Si_{0\cdot1})O_3$. These were produced from $BaCO_3$, TiO_2, GeO_2 and SiO_2 in the usual way by prefiring at 1120°c for 3 hours, pressing to 5 tons in^{-2} with an organic binder, and finally firing at 1200°–1350°c for 3 hours, the temperature required for a maximum density being found by trial. A reduction of firing temperature was observed for the substituted materials, $BaTiO_3$ firing to a maximum density of 5·54 g cm^{-3} at 1340°c, $Ba(Ti_{0\cdot9}Ge_{0\cdot1})O_3$ to 5·48 g cm^{-3} at 1200°c and $Ba(Ti_{0\cdot9}Si_{0\cdot1})O_3$ to 5·56 g cm^{-3} at 1300°c. The optimum x-ray densities are 6·03, 6·10 and 5·98 g cm^{-3} respectively.

X-ray powder photographs of the ceramics revealed a single tetragonal phase present in all three cases. These x-ray lines were somewhat diffuse for all three materials. This can perhaps be attributed to the crystallite size in the fired ceramics. No shift of high-angle reflections could be detected, indicating an unchanged unit cell size in the substituted materials. These results suggest that in substitution Si and Ge are analogous to Zr and Sn in occupying Ti sites, surrounded by an octahedron of O ions. This is noteworthy, since in pure silicates and germanates O ions are tetrahedrally disposed round the quadrivalent ion, whereas in $BaZrO_3$ and $BaSnO_3$ (which have the same perovskite structure as $BaTiO_3$) the O ions are octahedrally disposed.

The dependence of small-signal permittivity on temperature for the three materials is shown in the figure. It may be noted that the principal effect is a reduction of the sharpness of the permittivity maxima at the transitions at 120°c and 10°c, without displacing them appreciably to higher or lower temperatures. The fact that the two substituted materials have curves lying close to each other indicates that the effects of Si and Ge are about equivalent, ion for ion.



The effect on permittivity of replacement of Ti in $BaTiO_3$ by Si and Ge.

Hysteresis loops were measured at 50 c/s using a peak field of 28·3 kv cm^{-1}, and values of remanent polarization, coercive field and 'total' permittivity (slope of the line joining the extremities of the loop) were measured. Their values are given in the accompanying table.

Material	$BaTiO_3$	$Ba(Ti_{0.9}Ge_{0.1})O_3$	$Ba(Ti_{0.9}Si_{0.1})O_3$
Remanent polarization (μc cm^{-2})	5·5	4·6	1·3
Coercive field (kv cm^{-1})	7·4	5·5	3·4
Total permittivity	5540	3160	3550

ACKNOWLEDGMENTS

We are indebted to Dr. R. M. Hinde for the x-ray powder photographs, and to Dr. J. B. Birks, Research Manager of British Dielectric Research Ltd., for permission to publish this note.

REFERENCES

COFFEEN, W. W., 1953, *J. Amer. Ceram. Soc.*, **36**, 215.

DUNGEN, R. H., KANE, D. F., and BICKFORD, L. R., 1952, *J. Amer. Ceram. Soc.*, **35**, 318.

NATIONAL LEAD COMPANY OF AMERICA, Titanium Alloy Manufacturing Division, 1951, *High Dielectric Constant Titania and Titanate Ceramics.*

SMOLENSKIĬ, G. A., TARUTIN, N. P., and GRUDTSIN, N. P., 1954, *Zh. Tekh. Fiz.*, **24**, 1584.

Surface States on Germanium

BY J. LEES AND S. WALTON
Research Laboratories of The General Electric Company Ltd., Wembley, Middlesex

MS. received 22nd March 1955, *and in amended form* 13*th July* 1955

THE direct current voltage characteristics of tungsten contacts on various resistivities of p- and n-type germanium have been obtained. The tungsten was in the form of a 200 μ diameter whisker, etched electrolytically at the tip to a 150° cone. A load of 15 g was applied to each contact. The germanium specimens, cut from a single crystal ingot in which a large range of p- and n-type resistivities occurred, were etched in CP4 (Haynes and Shockley 1951). The sign of the majority carrier in each specimen was determined by measurements of the Hall effect.

The electrical characteristics are shown in figure 1. There is a definite trend as we pass from low resistivity p-type, through intrinsic, to low resistivity n-type material. On 1·5Ω cm p-type germanium there is rectification in the p-type direction. Slight p-type rectification also occurs on the 6·9 Ω cm sample, and on the 25 Ω cm specimen the characteristics are symmetrical. It is probable that heating at the contact accounts to a large extent for the small departures from linearity of the characteristics on these two samples. The 40 and 60 Ω cm p-type specimens show rectification in the direction corresponding to n-type material. The n-type specimens all show marked n-type rectification.

These results are in agreement with the surface states theory due to Bardeen (1947). At the surface of a semiconductor such as silicon or germanium additional energy levels exist; it is assumed that there is zero net charge in the levels when they are occupied by electrons to some level in the forbidden gap lying a height ϵ_0 above the valence band at the surface. If the height ζ of the Fermi level above the valence band in the interior of the semiconductor is greater than ϵ_0, a potential barrier to electrons is formed at the surface when equilibrium is established between the surface and the interior. If ζ is less than ϵ_0 a barrier to holes is formed and if ζ equals ϵ_0 there is no potential barrier at the surface. When the surface state density is large, the band structure at the semiconductor surface is unaffected by the presence of a contacting metal.

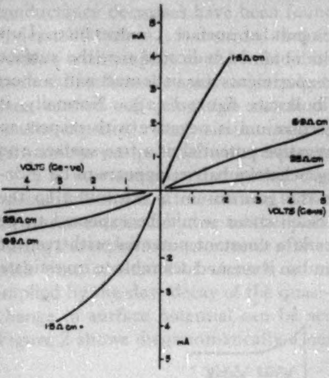


Figure 1 (*a*). Tungsten–p-type germanium I/V characteristics.

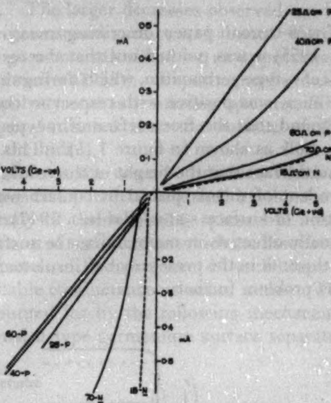


Figure 1 (*b*). Tungsten–germanium I/V characteristics.

On the basis of this model an estimate of ϵ_0 may be made from the observed electrical characteristics. The lowest degree of rectification is observed on 25 Ω cm p-type material; thus for this resistivity ζ is approximately equal to ϵ_0. ζ may be calculated from the known mobilities μ_n and μ_p of electrons and holes in germanium, combined with the relation between the densities of electrons and holes observed by Morin and Maita (1954) for germanium in the intrinsic range. Assuming an effective mass for holes of 0·3 m_e (Benedict 1953), ζ for 25 Ω cm material is 0·27 ev. Thus we conclude that ϵ_0 lies about 0·27 ev above the valence band.

For high resistivity p-type germanium, ζ is greater than ϵ_0 and a barrier to electrons, similar to that on n-type material, is to be expected, although in this case the compensating positive space charge is due to holes. Such a barrier may account for the n-type rectification observed on the specimens of high p-type resistivity, since an appreciable proportion of their current carriers are electrons.

References

Bardeen, J., 1947, *Phys. Rev.*, **71,** 717.
Benedict, T. S., 1953, *Phys. Rev.*, **91,** 1565.
Haynes, J. R., and Shockley, W., 1951, *Phys. Rev.*, **81,** 835.
Morin, F. J., and Maita, J. P., 1954, *Phys. Rev.*, **94,** 1525.

Modulation of the Surface Conductance of Germanium by Pulsed Electric Fields

By G. G. E. LOW
Department of Physics, University of Reading

Communicated by R. W. Ditchburn ; MS. received 9th August 1955

In a recent paper describing measurements of surface conductance (Low 1955) it was pointed out that the results obtained indicated that the surface of n-type germanium, which during the experiments was in contact with a sheet of mica, was positive with respect to the bulk (see figure 1 (*a*)). Normally, it is found that the free surface of n-type germanium is negative with respect to the bulk as shown in figure 1 (*b*). This negative potential at a free surface and also the fact that the height of the resulting Schottky barrier appears to be independent of metals placed in contact with the germanium is attributed to the action of surface states (Bardeen 1947). Such states would be expected to be equally effective in maintaining the surface at a constant potential with respect to the bulk in the presence of an insulator, and so it seemed desirable to investigate this problem further.

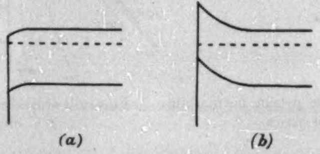


Figure 1. Potential energy profiles at the surface of n-type germanium for (*a*) surface positive with respect to the bulk, and (*b*) surface negative with respect to the bulk.

The original experiments consisted of measuring the conductance changes observed subsequent to the application of an external electric field to the germanium surface. The field was applied by means of a metal plate placed parallel to the germanium surface and separated from it by a mica spacer. By applying rectangular voltage pulses to this condenser-like system and displaying the resulting conductance variations on an oscilloscope it was observed that the initial conductance changes decayed to quasi-stable values with time constants of tens of microseconds. In addition it was found that the quasi-stable conductance values decayed further to zero, but owing to the long time constants involved (seconds or minutes) the pulse method could not be used to observe this process and approximate measurements only were made by means of an ordinary Wheatstone bridge circuit and galvanometer. The results, indicating that the surface of n-type germanium was positive with respect to the bulk, were obtained by applying negative pulses to the metal plate and germanium system. Under these conditions the quasi-stable conductance values observed were less than the original equilibrium conductance. As the induced charge was positive, this was consistent with the loss of mobile electrons and indicated that these carriers predominated in surface conductance. The magnitude of the decreases in conductance, however, were such that they could be accounted for only by assuming that the concentration of electrons in the conduction band increased close to the surface, i.e. that the surface was positive with respect to the bulk.

Further experiments of this type have been carried out and the results have confirmed the original observations. The use of a high voltage pulse generator has enabled greater fields to be applied and considerable quasi-stable conductance decreases have been found. The larger decreases observed can be understood only if it is assumed that the germanium surface was about 0·06 volt positive with respect to the bulk (this figure was calculated by an approximate method based on the analysis given by Schrieffer (1955)). Experiments have also been carried out using thin sheets of polystyrene as insulating spacers. The results obtained were similar and again indicated that the surface of the germanium was positive with respect to the bulk.

It is now considered that this change from the negative potential which is normally found at a free surface to the positive potential indicated by the above measurements is not due to the proximity of the insulating spacer but arises during the experiment as a result of the action of surface traps which are associated with time constants of seconds or minutes. (The presence of such traps is implied by the slow decay of the quasi-stable conductance values to zero.) The change in surface potential can be accounted for by the following mechanism. Figure 2 shows diagrammatically a normal n-type germanium surface separated by an insulator from a metal plate. This system may be roughly represented by the electrical circuit shown beside it. The condenser C_1 corresponds to the capacity between the metal plate and the surface traps, while C_2 represents the electrostatic interaction between the traps and the germanium. (For the sake of simplicity only one set of traps will be considered.) The resistance R gives a measure of the capture cross section of the traps, and consequently the product RC_2 has a value of seconds or minutes. Hence, as the repetition frequency of the pulses in the present experiments was 50 c/s only a small fractional change in the charge on the part of the circuit representing the surface traps can take place between pulses, and during an experiment this portion of the circuit will acquire an almost constant charge. Calculation shows that this charge is approximately given by C_1Vf where V is the voltage of the pulses and f is the duty cycle, i.e. pulse length times repetition frequency. This expression is not

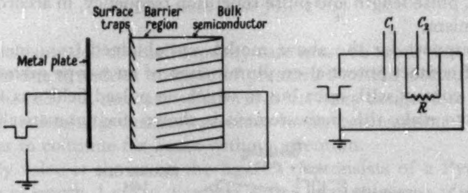


Figure 2. Diagram showing arrangement of metal plate and germanium during an experiment and approximate equivalent circuit discussed in text.

expected to apply at all accurately to the actual metal plate and germanium system as (*a*) the condenser C_2 does not represent the relation between potential and charge when the effect of space charge in the germanium is taken into account, and (*b*) the relation between the potential of the surface traps and their rate of loss or gain of charge is by no means linear as is implied by the use of the resistance R. The error introduced by approximation (*b*) is expected to be large compared with that due to (*a*), and arguments involving trapping rates indicate that the elimination of approximation (*b*) would lead to a mean trap charge considerably exceeding C_1Vf. It is believed that this charge may well be sufficient to produce the change in surface potential necessary to account for the large quasi-stable conductance decreases observed.

Confirmation that the surface traps do charge up in this manner during an experiment was obtained by measuring the mean change in resistance (by the use of an ordinary moving-coil ohm-meter) of the specimen when pulses were applied. For high voltage negative pulses a mean decrease in resistance was found corresponding to an increase in surface conductance of about 7×10^{-5} ohm^{-1}. Calculation shows that the effect of the decreases in conductance which result during each pulse is negligible in comparison with this figure, so that this increase may be taken to be the result of a change in surface potential arising from positive charge induced in the traps. An approximate calculation shows that this change corresponds to the surface potential being increased from its usual negative value (figure 1 (*b*)) to 0·07 volt positive with respect to the bulk, which makes it possible to interpret the results described above. The mean change in surface conductance measured in this way was found to increase with pulse voltage, pulse length and pulse repetition frequency, in accordance with the above mechanism.

Further support for the above model was obtained from measurements of the change of contact potential on illumination of an n-type germanium surface which was in contact with mica but to which no pulsed field was applied. The method used to make this measurement is shown diagrammatically in figure 3.

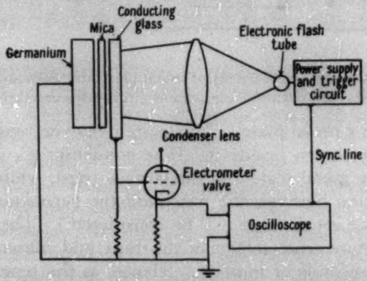


Figure 3. Schematic diagram of the method used to measure the change of contact potential on illumination.

Short pulses of light from an electronic flash tube were focused by means of a condenser lens through a sheet of conducting glass and the insulating spacer on to the germanium surface. The changes of contact potential resulting from the light pulses gave rise to e.m.f's which appeared on the grid of the electrometer pre-amplifier and hence were recorded by the oscilloscope. The results obtained were of the same sign and order of magnitude as the signals observed when measurements were carried out by the same method on a free germanium surface, the sign of the effect in both cases being consistent with a barrier of the type shown in figure 1 (*b*).

The above discussion is in terms of positive induced charge on germanium but it is clear that the mechanism described will also operate when the induced charge is negative.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for placing research facilities at his disposal, Drs. P. C. Banbury and H. K. Henisch for valuable discussions, and the New Zealand Defence Research Organization for financial support.

REFERENCES

BARDEEN, J., 1947, *Phys. Rev.*, **71**, 717.
LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 10.
SCHRIEFFER, J. R., 1955, *Phys. Rev.*, **97**, 641.

A Mercury Safety Valve for Electrolysis of Heavy Water

BY R. M. CHAUDHRI
Physics Department, Government College, Lahore, Pakistan

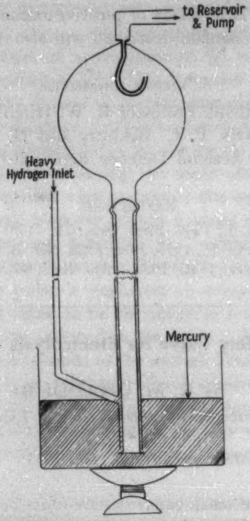
MS. received 26th July 1955

In an electrolysing cell of small capacity care must be taken that the electrolyte is not split due to sudden changes of pressure over the electrolyte in the two sides of the cell. A mercury safety valve has been designed to minimize this difficulty. It maintains a constant pressure in the cell enabling the electrolysis of heavy water to continue for hours without attention.

The safety valve is shown in the figure. It consists of a Pyrex glass tube about 100 cm in length, 1 cm in diameter with a wall thickness of about 1·5 mm. Its lower end dips into a mercury cup resting on an adjustable stand. A second Pyrex tube of slightly smaller diameter is slipped inside it leaving a narrow annular space between the two. The lower end of the inner tube is closed and flattened making it wider than the bore of the outer one. It is held in position by a bulb blown with an uneven surface at its upper end and sitting in a cavity in the outer tube. The bottom flat end and the bulb prevent it from moving up or down to any appreciable extent. The outer tube is closed at the upper end by a Pyrex glass flask of about 500 cm^3 capacity and a hooked Pyrex glass capillary 1 mm in bore at its top connects it to the pumping side and the reservoir.

A narrow Pyrex glass tube sealed on one side of the outer tube serves as an inlet for heavy hydrogen.

As the tube is exhausted mercury rises in the annular space to a height equivalent to that of the atmosphere and its level in the cup is adjusted so that it just covers the side joint of the inlet tube. Heavy hydrogen produced by electrolysis bubbles through mercury against the constant atmospheric pressure while oxygen escapes to the atmosphere. The bubbling of heavy hydrogen causes the splashing of mercury drops which may enter the flask through the narrow openings left between the inner and outer tubes but they will have very little chance to enter the glass capillary hook and pass on to the reservoir. Heavy hydrogen, however, flows freely to it.



As hydrogen has to bubble into a vacuum, it may overshoot and disturb the equilibrium pressure of the cell which would allow air to rush in through the oxygen delivery tube, if it is open to the atmosphere. It is, therefore, just dipped in a cup of mercury which acts as a valve preventing air rushing in but allowing oxygen to escape.

My thanks are due to Mr. Abdul Latif Taseer for preparing the figure and Mr. Ghulam Safdar for assembling the heavy hydrogen plant.

CORRIGENDUM

On the Jet Formation by Explosives with lined Hemispherical Cavities, by SAMPOORAN SINGH (*Proc. Phys. Soc.* B, 1955, **68**, 785).

On page 788 the sentence before equation (6) should read: Let the cylindrical coordinates of R be (r, z) and the coordinates of P$'$, the original position of R in the liner, be $(\frac{1}{2}D \cos \alpha, M')$ where M' is defined by $T=M'/U_{\mathrm{d}}$.

In equation (9), the third term in the last bracket should be $V_0' \tan A/V_0$.

REVIEWS OF BOOKS

Siebenstellige Tafeln der elementaren transzendenten Funktionen, by F. Lösch. Pp. viii + 335. (Berlin – Göttingen – Heidelberg: Springer-Verlag, 1954.) DM. 49.80 (87*s.* 2*d.*)

The functions in tables I and II are: angle ϕ, $\sin x$, $\cos x$, $\tan x$, $\sinh x$, $\cosh x$, $\tanh x$, $\ln x$, e^x, e^{-x}, $\sin^{-1} x$, $\tan^{-1} x$, $\sinh^{-1} x$, $\cosh^{-1} x$, $\tanh^{-1} x$, $\coth^{-1} x$, where these functions are real. In both tables ϕ is given in degrees, minutes, seconds and 2 D of a second. Table I gives the remaining functions to 9 D for $x = 0(0{\cdot}0001)0{\cdot}1$; table II gives them to 7 D for $x = 0{\cdot}1(0{\cdot}0005)3{\cdot}15$, $x = 3(0{\cdot}01)10$, $x = 10(0{\cdot}1)20$ together with $\tan x$ for $x = 1{\cdot}5680(0{\cdot}0001)1{\cdot}5730$, $\tanh^{-1} x$ for $x = 0{\cdot}9980(0{\cdot}0001)1$ and $\coth^{-1} x$ for $x = 1(0{\cdot}0001)1{\cdot}0020$. Table III gives several of the same functions as I and II, to 7 D for $x = 0(1)100$. Table IV gives $\frac{1}{2}x\pi$ to 12 D for $x = 0(1)100$. Table V gives $\sin \frac{1}{2}x\pi$, $\cos \frac{1}{2}x\pi$ to 7 D for $x = 0(0{\cdot}001)0{\cdot}5$. Table VI gives $\exp(\pm \frac{1}{2}x\pi)$, $\sinh \frac{1}{2}x\pi$, $\cosh \frac{1}{2}x\pi$ to 7 D for $x = 0(0{\cdot}01)2$; table VII gives the same functions of $x\pi/180$ to 7 D for $x = 0(1)180$. Table VIII gives angle in degrees, minutes and seconds and also in degrees and decimals of a degree, as a function of angle in radians; table IX is the inverse of table VIII. Table X gives about 20 important constants and (as exact ratios) Bernouilli numbers as far as B_{40} (B_{20} in Glaisher's notation); factorials as far as 15 ! are also given.

English readers will naturally compare these tables with the second volume of Comrie's 1949 edition of *Chambers' Six-Figure Mathematical Tables*; these are six-figure or six-decimal, whichever is the less, whilst Lösch's gives not less than seven decimals. Again, Comrie's are tabulated only in the first quadrant, for the circular functions. On the other hand, Lösch's tables omit the reciprocal functions altogether and do not tabulate the circular functions against angle in degrees. Lösch's notes on interpolation are meagre in the extreme, not going beyond the ill-conceived recommendation of the Newton–Gregory forward formula. The contrast with Comrie's masterly introduction is as striking as the contrast in price. First differences are given where interpolation is supposed to be practicable and it is not perhaps surprising that the author gives up as soon as Δ''' becomes significant. No guidance is given to the reader as to how to enter such a table as e^x at $x \sim 5$, say, where, of course, special methods are available.

According to the preface, the *raison d'etre* of the tables is that Hayashi's tables (Berlin : Springer 1926) are no longer available. It might have been added that they reached a level of unreliability rivalled only by the extensive "corrections" issued by Hayashi himself in 1932. We are told, however, that the present tables have been abbreviated from ten-decimal values, which had been checked by differencing. The quality of the paper is good and the type is clear, except that the italicized differences leave something to be desired in this respect.

J. C. E. JENNINGS.

CONTENTS OF SECTION A

PAGE

Mr. W. HAYES. The Spectra of the Manganese Halides in the Blue-Green . 1097

Mr. W. R. S. GARTON and Mr. A. RAJARATNAM. On the Levels of the p^2-Configuration in the Arc Spectra of Zn, Cd and Hg 1107

Dr. P. G. KLEMENS. The Scattering of Low-Frequency Lattice Waves by Static Imperfections 1113

Prof. C. A. COULSON. The Free-Electron Network Model for Metals : II . 1129

Mr. H. N. V. TEMPERLEY. The Liquid ' Cell ' Model Applied to ^{3}He . . 1136

Dr. N. H. MARCH. Equations of State of Elements from the Thomas–Fermi Theory : II—Case of Incomplete Degeneracy 1145

Mr. A. P. STONE. Calculation of Isotope Shifts in the Spectrum of Helium . 1152

Dr. P. SWAN. The Ionization by Electrons of the Excited 2s and 2p States of Atomic Hydrogen 1157

Prof. G. S. RUSHBROOKE and Mr. P. J. WOOD. On the High-Temperature Susceptibility for the Heisenberg Model of a Ferromagnet 1161

Mr. J. R. COOK. Elastic Scattering of Gamma Rays 1170

Mr. D. G. E. MARTIN and Dr. G. PARRY. A Search for Low Intensity Gamma Rays in the ThC . C′ and ThC″ . D Transitions 1177

Dr. P. G. HARPER and Dr. J. J. O'DWYER. Theory of a Dielectric Loss in Long Chain Polar Compounds 1184

Research Notes :

Prof. H. JONES. The Thermoelectric Power of Monovalent Metals . . 1191

Mr. S. N. BISWAS. Many-Body Scattering Problem in Momentum Space 1193

Mr. N. W. TANNER. Angular Distribution of γ-Radiation from $^9Be(\alpha, n\gamma)^{12}C$ 1195

Mr. A. WEINMANN. Inequalities for Molecular Integrals 1197

Prof. D. R. BATES and Mr. T. R. CARSON. Doubly Charged Diatomic Molecular Ions 1199

Reviews of Books 1203

Contents of Section B 1208

Subject Index, Section A, Vol. **68** 1209

Index of Authors (with Titles), Section A, Vol. **68** 1217

Index to Reviews of Books, Section A, Vol. **68** 1227

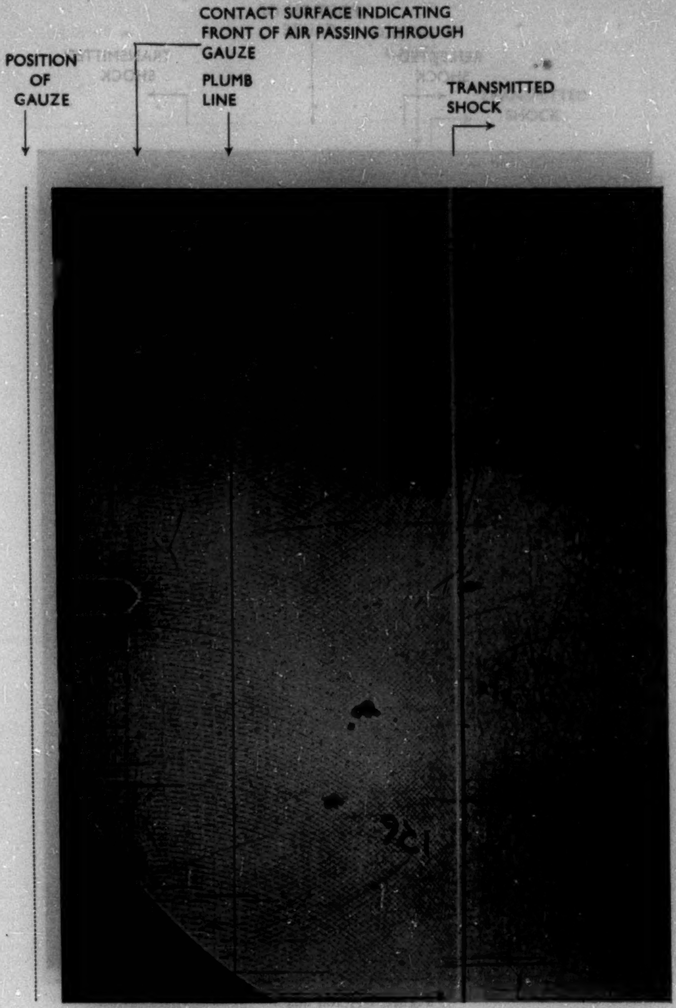

POSITION OF GAUZE
CONTACT SURFACE INDICATING FRONT OF AIR PASSING THROUGH GAUZE
PLUMB LINE
TRANSMITTED SHOCK

Figure 2 ($\frac{3}{4}$ full size).

PLATE I.

PROC. PHYS. SOC. VOL. 68, PT. 12—B (J. E. BOWMAN AND G. B. F. NIBLETT)

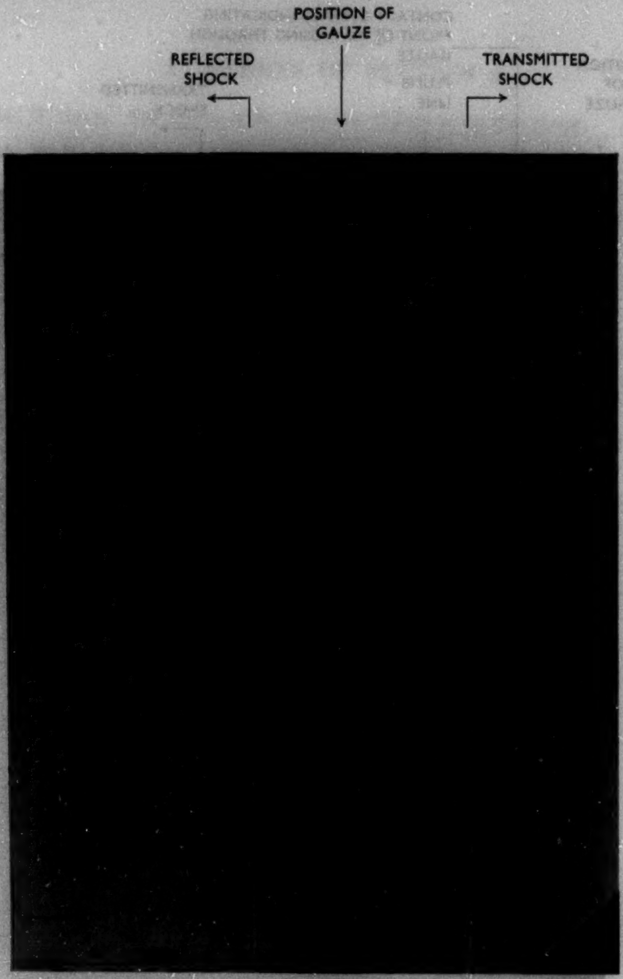

POSITION OF GAUZE
REFLECTED SHOCK
TRANSMITTED SHOCK

Figure 4 (¾ full size).

Plate II.

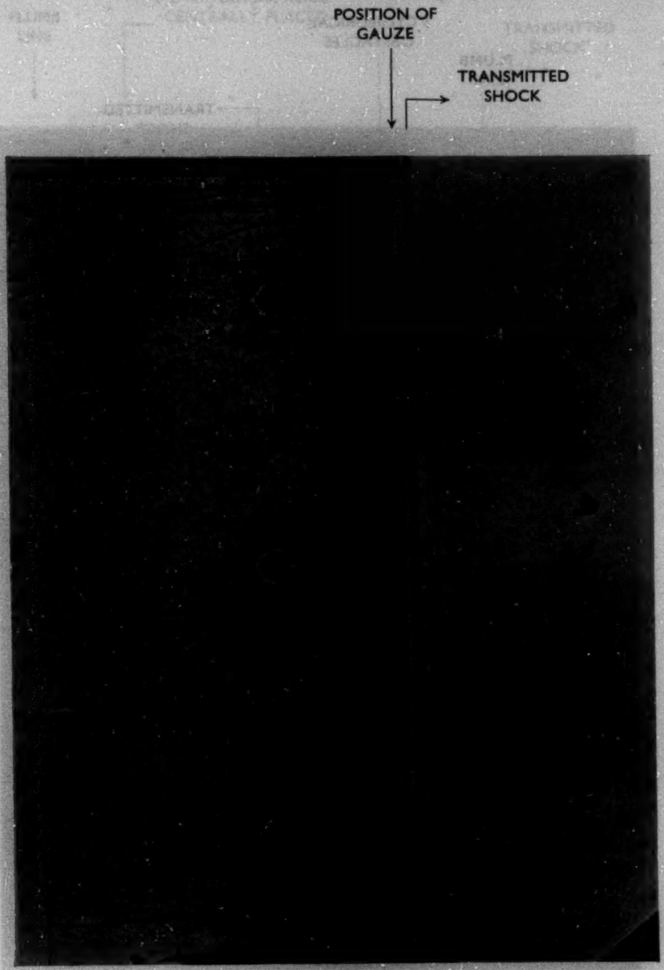

POSITION OF GAUZE
TRANSMITTED SHOCK

Figure 6 ($\frac{3}{4}$ full size).

PLATE III.

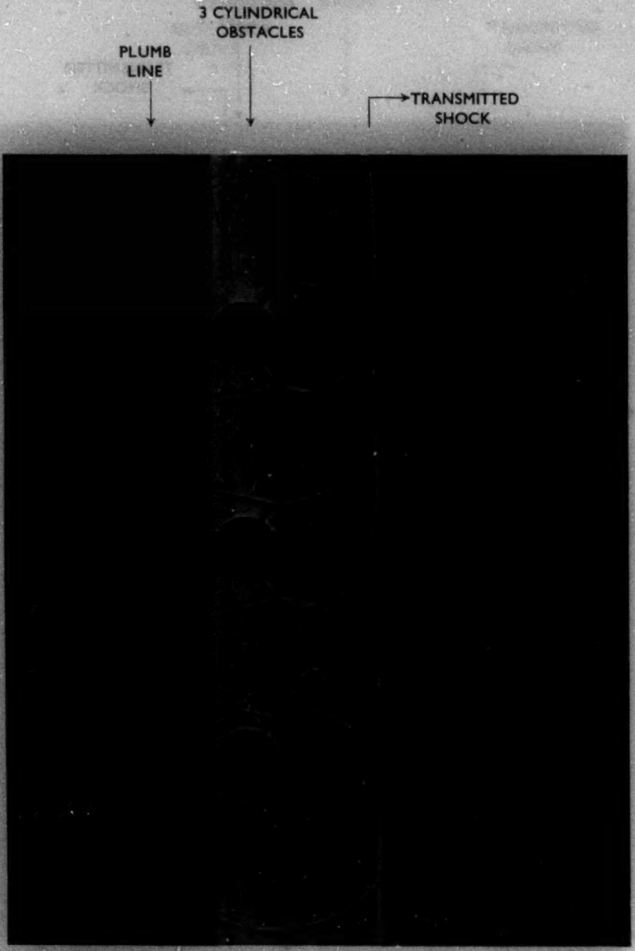

3 CYLINDRICAL OBSTACLES
PLUMB LINE
TRANSMITTED SHOCK

Figure 7 ($\frac{3}{4}$ full size).

PLATE IV.

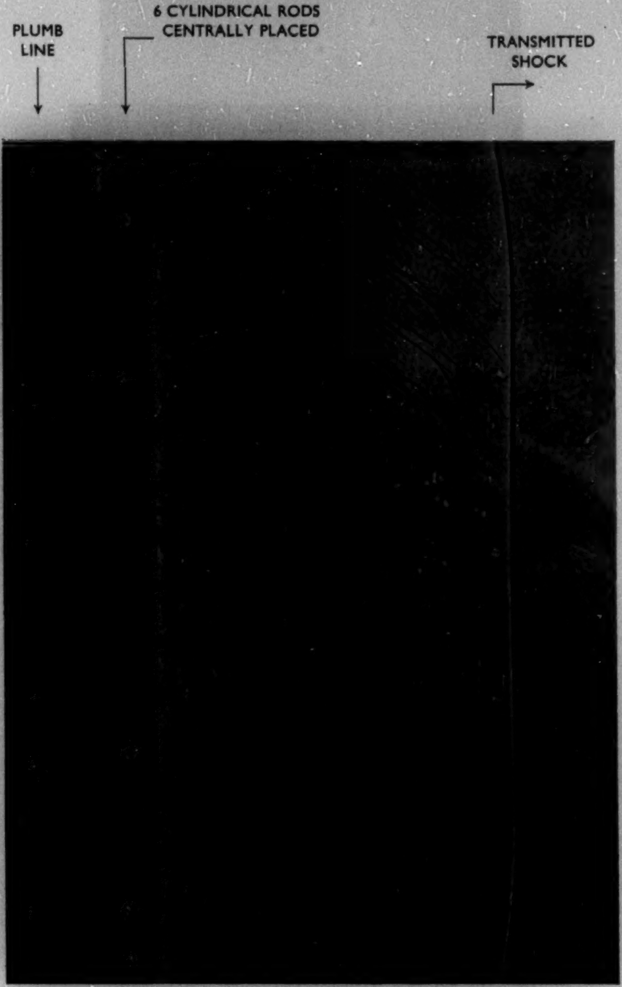

PLUMB LINE
6 CYLINDRICAL RODS CENTRALLY PLACED
TRANSMITTED SHOCK

Figure 8 ($\frac{3}{4}$ full size).

Plate V.

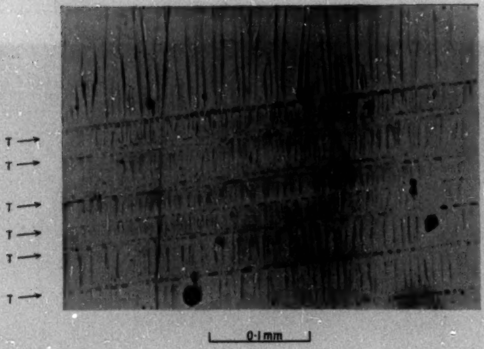


Figure 2. General view of domain pattern (T indicates a twin band).

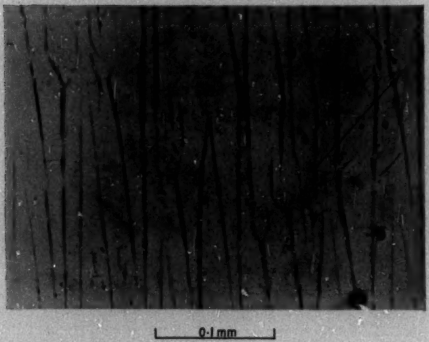


Figure 3. Domain pattern near a twin-free fracture edge.

PLATE I.

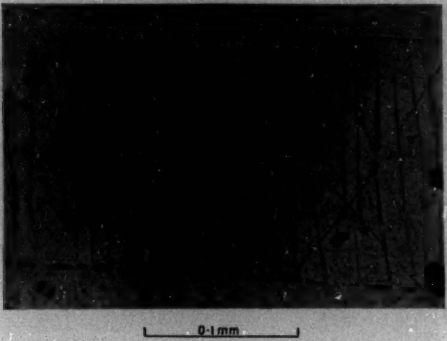


Figure 4. Domain pattern with applied field zero.

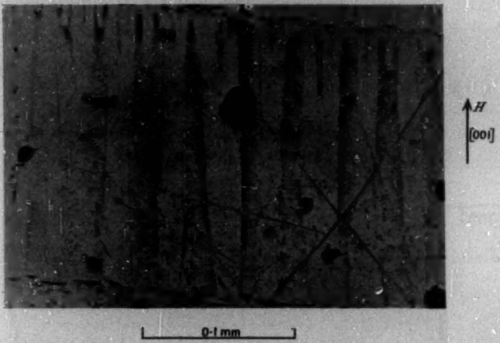


Figure 5. Domain pattern with field parallel to [001]. $H \simeq 20$ Oe.

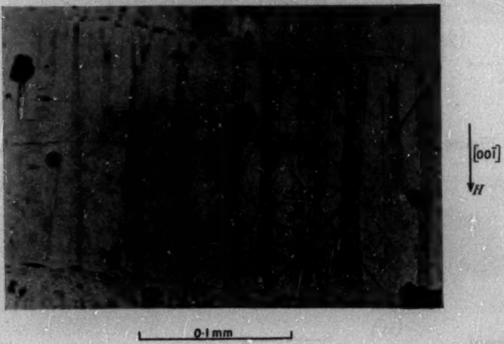

[00$\bar{1}$]
H
0·1mm

Figure 6. Domain pattern with field reversed parallel to [001]. $H \simeq 20$ Oe.

PLATE II.

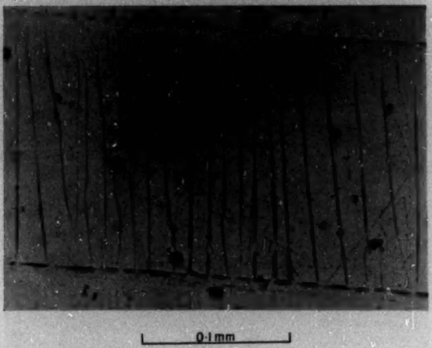


Figure 7. Domain pattern with zero field.

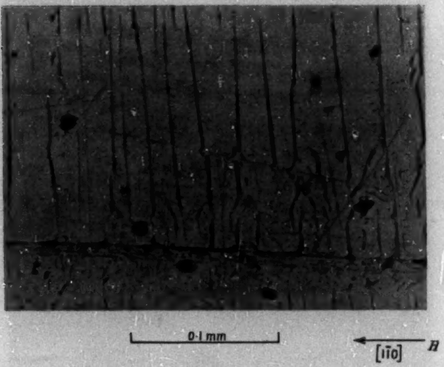

0·1 mm
$[1\bar{1}0]$
H

Figure 8. Domain pattern with field parallel to $[1\bar{1}0]$. $H \simeq 20$ Oe.

PLATE III.

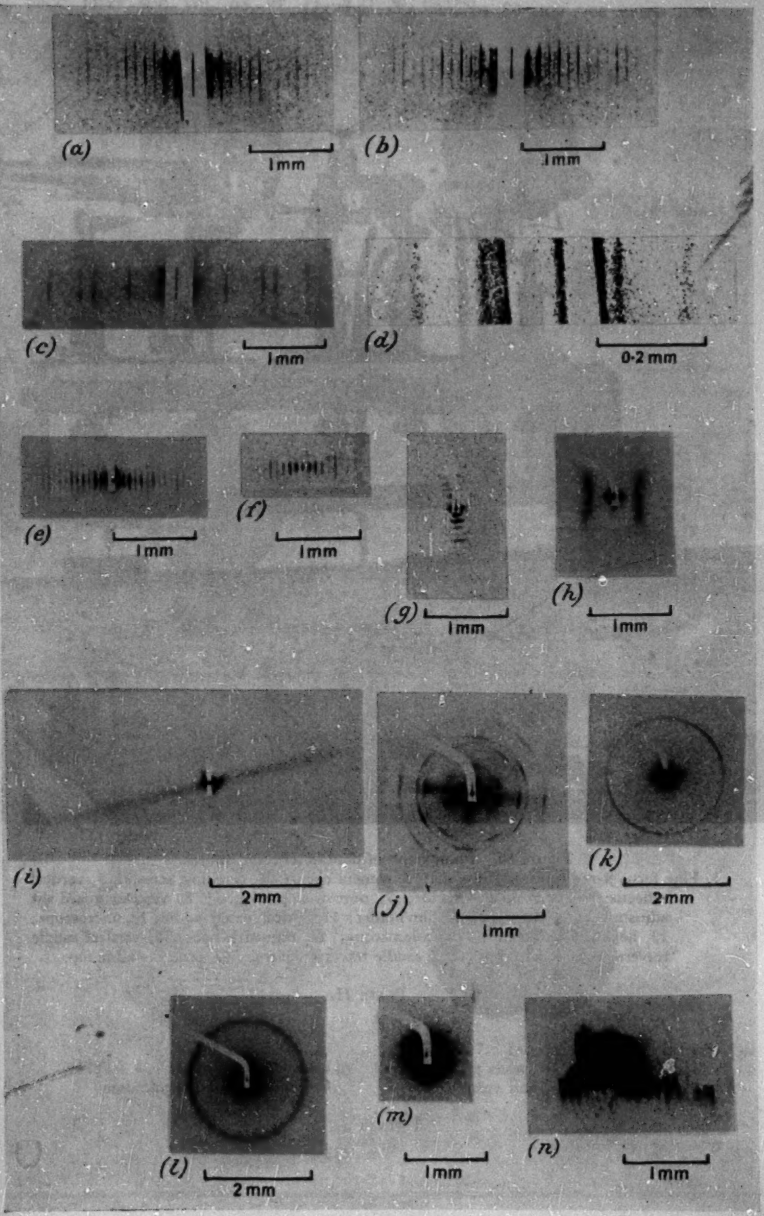

(a)
1mm
(b)
1mm
(c)
1mm
(d)
0·2 mm
(e)
1mm
(f)
1mm
(g)
1mm
(h)
1mm
(i)
2mm
(j)
1mm
(k)
2mm
(l)
2mm
(m)
1mm
(n)
1mm

Figure 4.

Plate I.

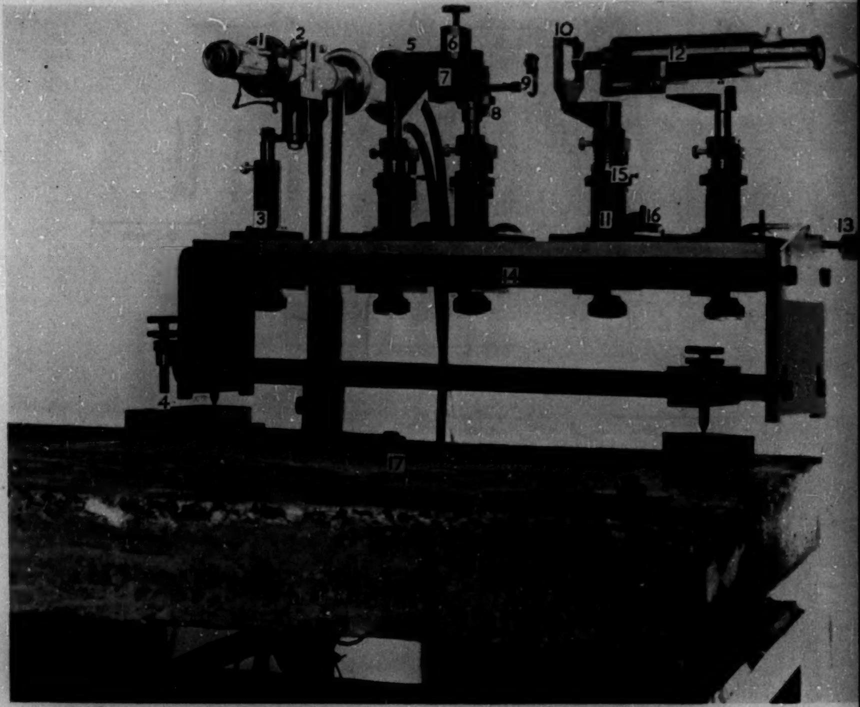


Figure 10. Photograph of double reflector camera.

1, Fine focus x-ray tube; 2, tube slit; 3, camera clamp; 4, levelling screws; 5, vertical reflector; 6, horizontal reflector; 7, mirror stop control; 8, vertical guard slit adjustment; 9, guard slit; 10, film holder; 11, optical bench saddle; 12, microscope; 13, horizontal adjustment for microscope; 14, optical bench; 15, vertical saddle traverse control; 16, horizontal saddle traverse control; 17, concrete table top.

PLATE II.

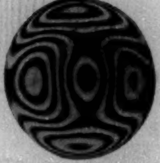


Figure 2. Twyman interferogram: 20 in. Aviar lens f/5·6.

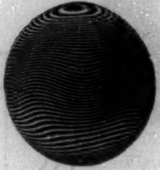


Figure 3. Same as figure 2, but with a tilt of the reference mirror.

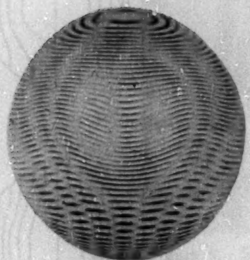


Figure 4. Moiré pattern formed by combining two copies of figure 3.

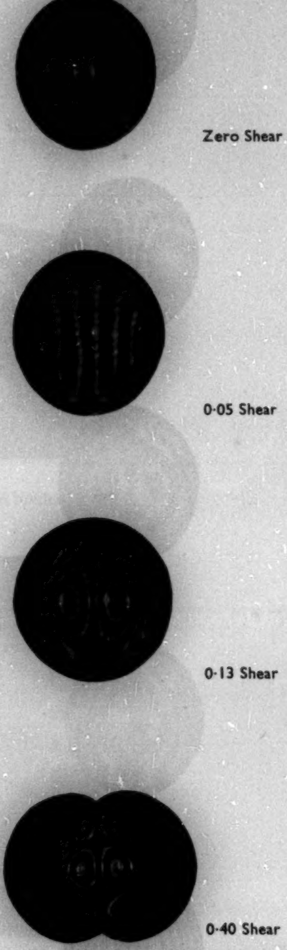
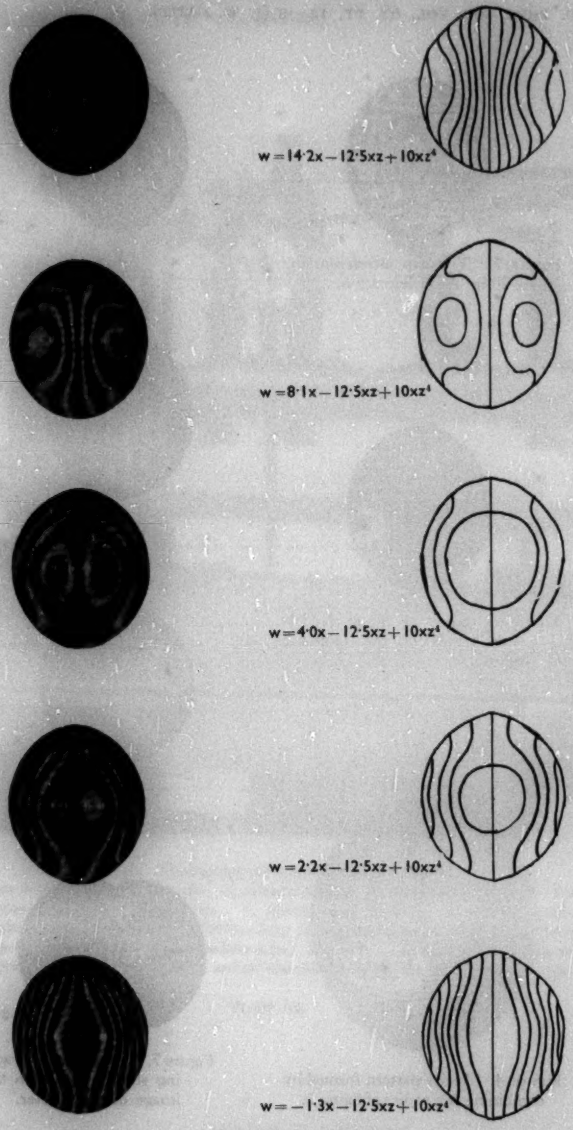

Zero Shear
0·05 Shear
0·13 Shear
0·40 Shear

Figure 7. Interferograms with increasing shear, made with the double-image interferometer.

PLATE I.



$w = 14{\cdot}2x - 12{\cdot}5xz + 10xz^4$

$w = 8{\cdot}1x - 12{\cdot}5xz + 10xz^4$

$w = 4{\cdot}0x - 12{\cdot}5xz + 10xz^4$

$w = 2{\cdot}2x - 12{\cdot}5xz + 10xz^4$

$w = -1{\cdot}3x - 12{\cdot}5xz + 10xz^4$

Figure 6. Interferograms from the double-image interferometer, compared with the patterns obtained by plotting the sum of the three most important aberration terms (20 in. Aviar lens, f/5·6, as in figures 2, 3 and 4).

PLATE II.

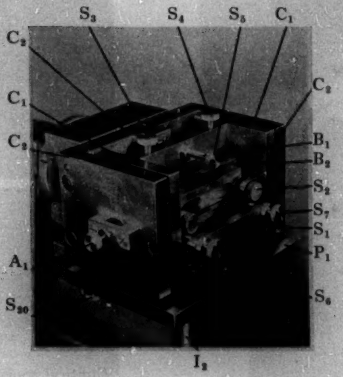

S_3
S_4
S_5
C_1
C_2
C_2
C_1
C_2
B_1
B_2
S_2
S_7
S_1
P_1
A_1
S_6
S_{20}
I_2

Figure 4. The crystal holder.

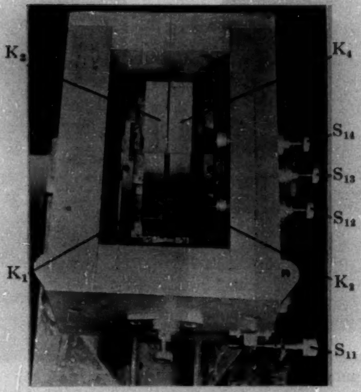
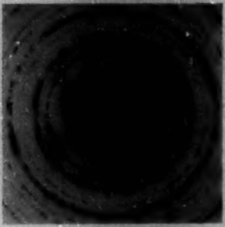

K_2
K_4
S_{14}
S_{13}
S_{12}
K_1
K_3
S_{11}

Figure 9. Source holder without lid.

PROC. PHYS. SOC. VOL. 68, PT. 12—B (P. B. HIRSCH, A. KELLY AND J. W. MENTER)



2 (*a*)

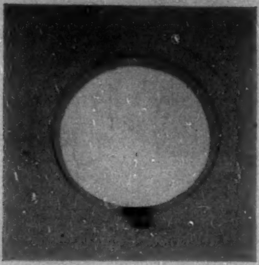
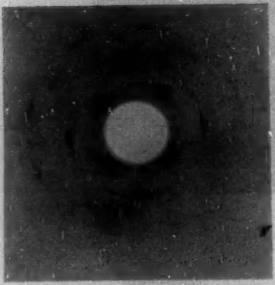
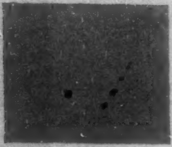


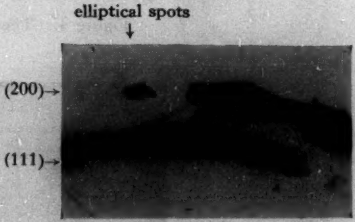
Figure 1. X-ray micro-beam (8 μ diameter) transmission photograph of beaten gold foil.



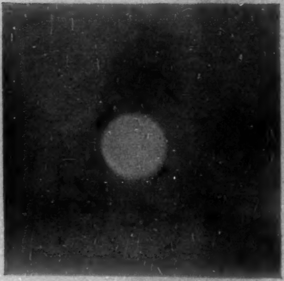
2 (*b*)



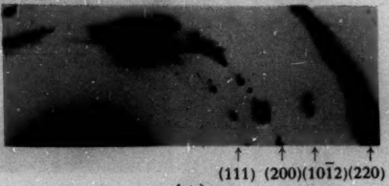
4 (*a*)



4 (*b*)



2 (*c*)

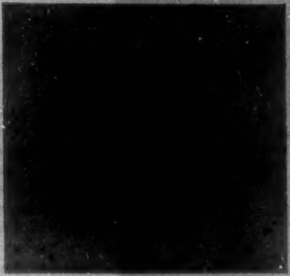


4 (*c*)

Figure 2. Electron diffraction transmission patterns of beaten gold foil. (*a*) Beam diameter 10 μ. Typical 'random' orientation. Streaks can be seen between the (111) and (200) rings, and extra spots, from a hexagonal phase, can be seen between the (200) and (220), and the (220) and (311) rings respectively. (*b*) Beam diameter 10 μ. Strong (100) preferred orientation. (*c*) Beam diameter 2 μ. Different orientation. The arrangement of spots on opposite ends of the diameters on the (111) ring is nearly identical.

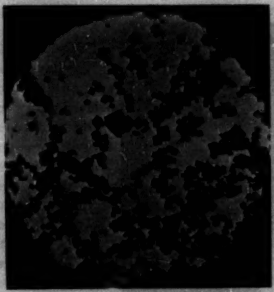
Figure 4. (*a*) Elliptical spots; the orientation of the ellipses is constant around the rings. (*b*) Elliptical spots marked with arrow; the orientation of the ellipses is variable. Streaks and displaced spots are clearly visible. (*c*) Streaks on (111) and (200) rings and elliptical spots on (10$\bar{1}$2).

PLATE I.

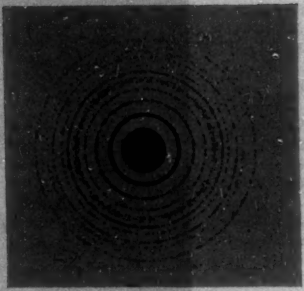


6 (*a*)

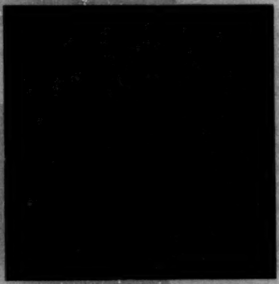
6 (*b*)



6 (*c*)



7 (*a*)



7 (*b*)

Figure 6. (*a*) Electron diffraction transmission pattern of magnesium oxide (beam diameter $10\,\mu$). (*b*) Enlargement of a diffraction ring showing clusters of spots due to refraction. (*c*) Electron micrograph of the same area (diameter $10\,\mu$) of the specimen as used for (*a*) ×7000.

Figure 7. (*a*) Electron diffraction pattern from an evaporated layer of tin (beam diameter $10\,\mu$). (*b*) Electron micrograph of the same area (diameter $10\,\mu$) of the specimen as used for (*a*) ×7000.

PLATE II.

Figure 2.

Figure 3.

PROCEEDINGS OF THE PHYSICAL SOCIETY

SECTION B, 1955—VOL. 68

SUBJECT INDEX

PAGE

Aberration, comatic, measurement by interferometry 1065
Aberration, spherical, in electron lenses, systems for correction 156
Absorption spectra, X-ray, *see under* X-ray.
Absorption spectrum of excited crystals of CdS. 758
Absorption, *see also* Dielectric.
Action-at-a-distance, theory 672
Adsorption of water vapour by anodized Al, electrical effects 1105
Air stream, circular cylinders in, measurement of sound 453
Alnico, domain structures in 537
Angle of contact, *see under* Contact.
Anthracene crystals, absolute quantum efficiency of photofluorescence . . . 241
Aromatic hydrocarbons, condensed, photoeffects in (R) 974
Atoms, thermal vibrations in cubic crystals *see under* Crystals.
Auger effect in semiconductors (L) 319
Auroral activity, diurnal variation 408
Automatic design of optical systems. 729

Barkhausen effect 1017
Becke line method of refractive index determination 1081
Binding energy of nitrogen in dislocations 1001
Boiling of liquids, delayed, and radio-carbon content (R) 686
Bolometer, superconducting, as detector of thermal radiation from low temperature
    sources 801
Borazole and boron nitride scintillators (L) 253

Calcium tungstate, photoconductivity in 148
Camera, x-ray diffraction, optically focusing 1054
Carbon resistor thermometers (R) 388
Carrier concentration disturbances in semiconductors. . . . . 310, (L) 793
Carrier deficit phenomena in Ge, measurements (R) 970
Carrier extraction in Ge 43
Carriers, added in InSb, lifetime 985
Carriers, minority, in semiconductors, lifetime measurement (R) 562
Cathodes, L-, d.c. and pulsed emission from (R) 1146
Cathodo-luminescence, field enhanced, in magnesium oxide (R) 566
Centimetric wavelengths, ionic conductivity of dilute KCl solutions 65
Ceramics, $BaTiO_3$, replacement of Ti by Si and Ge (R) 1150
Cerium ethyl sulphate, demagnetization experiments on 415
Cerium, metallic, magentic susceptibility 859
Circuit elements, temperature dependent, a.c. admittance 766
Coherence of light wave, measurement by Lloyd mirror (R) 1148
Cold work, effect on integrated X-ray intensities, and X-ray extinction . . . 577
Cold-worked gold, structure : I—study by electron diffraction 1132
Colour centres, radiation induced, in fused quartz 223
Conductance, surface, of Ge and Si, modulation by external electric fields . . 10
Conductivity, electrical, Hall coefficient, and magneto-resistance effect of lead
    sulphide and telluride 22
Conductivity, electron bombardment, effect of metal electrodes 620
Conductivity, *see also* Ionic.
Contact, angle, of mercury against steel, hysteresis of 297
Contact resistance and mechanical properties of surface films on metals . . . 625

PAGE
Contacts, metal–germanium, influence of pressure on 922
Contacts, metal–semiconductor, pulse method for measuring injection ratio 447
Contacts, point on Ge, measurement of injection ratio 17
Contacts, *see also* Germanium, semiconductors.
Cracked materials, apparent strength 723
Creep, transient, restoration, and recovery and recrystallization in metals 257
Crystals, anthracene, absolute scintillation efficiency 929
Crystal, cubic, internal conical refraction of transverse elastic waves in 81
Crystals, cubic, thermal vibrations of atoms in: I—x-ray scattering by lead single crystals 951
Crystals, cubic, thermal vibrations of atoms in: II—amplitude of atomic vibrations 957
Crystals, excited, of CdS, absorption spectrum 758
Crystals grown from impure unstirred melts, impurity distribution 393
Crystals, lead fluoride, growth from the melt 165
Crystals, NaCl monocrystals, scattering of light in 325
Crystals, organic, fluorescence excitation spectra and quantum efficiencies 701
Crystals, single, of silicon, lamellar defects in 111
Crystals, single, of tin and bismuth, directional hardness 548
Crystals, single, of ZnS, electroluminescence in 775
Crystal structure of lanthanum–strontium manganites 586
Crystals, twinned, of SiFe, magnetic domain structure in 1033
Crystals, *see also* Anthracene.

Deformation faults in cold-worked metals (R) 51
Deformation properties of friction junctions 609
Demagnetization experiments on cerium ethyl sulphate 415
Deposition, temperatures at surfaces of growing deposits and origin of stress in deposits (L) 474
Diamond, occasional mode of growth 559
Diamond, temperature dependence of dielectric constant (R) 315
Dielectric absorption in dilute liquid polar solutions 1038
Dielectric constant of aqueous ionic solutions 357
Dielectric constant of diamond, temperature dependence (R) 315
Dielectric constant of standard liquids, measurement 422
Dielectric, fluid, non-linear, electrostatic problem in 1
Dielectric layers, thin, electron bombardment effects 991
Dielectric properties of eugenol, glycerol and water, isothermal pressure dependence (R) 790
Dielectrics, *see also* Ceramics.
Differential analyser solution of water bells problem (L) 54
Diffraction of non-monochromatic electromagnetic wave by slit and grating 171
Diffraction, *see also* Electron diffraction, X-ray diffraction.
Diffusion in gas of electrons from small source (R) 381
Diffusion in moving striations (R) 383
Diffusion and mutual viscosity in liquid mixtures 209
Diodes, germanium, *see* Germanium.
Dipole moments, electric, computation (L) 186, corr. 253
Discharge, electrodeless, in hydrogen, growth 137
Discharge, hot-cathode, noise from 292
Discharge, *see also* Glow discharge, Negative glow.
Dislocation, binding energy of nitrogen in 1001
Dislocation, edge, Peierls–Nabarro model modification 1043
Domain structures in alnico 537
Domain structure, magnetic, in twinned crystals of SiFe 1033

Elastic waves, transverse, in cubic crystal, internal conical refraction 81
Electroluminescence in single crystals of ZnS 775
Electrolysis of heavy water, mercury safety valve for (R) 1157

PAGE
Electron avalanche, retarded by own space charge, growth . . . 369
Electron beams, low energy, anomalous emission effects . . . 805
Electron bombardment conductivity, effect of metal electrodes . . . 620
Electron bombardment effects in thin dielectric layers . . . 991
Electron diffraction study of structure of cold-worked gold . . . 1132
Electron emission, high-field, from tarnished metal surfaces . . . 504
Electron emission in spark gaps, interpretation of experiments . . . 513
Electron lenses, magnetic, asymmetrical, field distribution (R) . . . 679
Electron lenses, magnetic, pinhole (R) . . . 682
Electron lenses, magnetic, unified representation of properties . . . 737
Electron lenses, numerical ray tracing . . . 526
Electron lenses, possible chromatic correction system for . . . 817
Electron lenses, systems for correction of spherical aberration . . . 156
Electron microscopy, reflection, examination of oxide films by (R) . . . 178
Electron motion in gas in presence of magnetic field, experimental investigation . 334
Electrons, from small source, diffusion in gas (R) . . . 381
Electrostatic problem involving non-linear fluid dielectric . . . 1
Emission, d.c. and pulsed, from L-cathodes (R) . . . 1146
Emission, electron, *see* Electron.
Emission, field, effect on behaviour of semiconductor contacts . . . 881
Emission, field, and Fowler–Nordheim law, factors influencing . . . 938
Emission, field, from semiconductors . . . 746
Emission, secondary, *see* Secondary.
Entrainment and turbulence in open gas streams . . . 234
Equation of state of solids and interatomic force law . . . 1049

Faraday effect, infra-red, due to free carriers in semiconductor (R) . . . 973
Ferrites, heat changes accompanying magnetization processes . . . 304
Ferromagnetic resonance in Ni and alloys . . . 713
Field emission, *see* Emission.
Films, metallic, *see* Metallic films.
Films, oxide, *see* Oxide films.
Films, surface, on metals, contact resistance and mechanical properties . . . 625
Fluorescence excitation spectra and quantum efficiencies of organic crystals . . . 701
Fourier series, applications in numerical treatment of linear behaviour . . . 89
Fowler–Nordheim law and field emission, factors influencing . . . 938
Freezing of supercooled water . . . 193
Friction junctions, deformation properties . . . 609
Friction, relation to wear for boundary-lubricated surfaces . . . 603

Galvanomagnetic and thermoelectric effects in lead selenide and telluride . . . 35
Gas, electron diffusion in, with electrons from small source (R) . . . 381
Gas, electron motion in, in presence of magnetic field, experimental investigation . 334
Gas streams, open, entrainment and turbulence in . . . 234
Germanium, carrier deficit phenomena, measurements (R) . . . 970
Germanium, carrier extraction in . . . 43
Germanium diodes, illuminated, low frequency noise in (R) . . . 569
Germanium diodes, point-contact, barrier height inferred from measurements of voltage dependence of capacitance (R) . . . 113
Germanium, ionic bombardment of, effects produced by . . . 213
Germanium junctions, p–n alloy, breakdown effect in (R) . . . 182
Germanium, measurement of injection ratio of point contacts on . . . 17
Germanium, modulation of surface conductance by external electric fields . . . 10
Germanium, surface conductance, modulation by pulsed electric fields (R) . 1154
Germanium, surface recombination in strong electric fields . . . 353
Germanium, surface states on (R) . . . 1152
Glow discharge, constriction of positive column at high pressure . . . 130

PAGE
Gold, cold-worked, structure : I—study by electron diffraction 1132
Gravitation, Whitehead's law, in space–time of constant curvature 944
Growth, occasional mode, in diamond 559

Hall coefficient, electrical conductivity and magneto-resistance effect of lead sulphide, selenide and telluride 22
Halogens, solid, infra-red luminescence (L) 53
Hardness, directional, of single crystals of Sn and Bi 548
Heat changes accompanying magnetization processes in ferrites 304
Heavy water, electrolysis, mercury safety valve for (R) 1157
Helium temperatures, resistance thermometer for (R) 386

Illumination, effect on excess noise of Si filaments (R) 968
Impact of solid on liquid surface, sounds of 541
Impurity distribution in crystals grown from impure unstirred melts 393
Infra-red Faraday effect due to free carriers in semiconductor (R) 973
Infra-red luminescence of solid halogens (L) 53
Infra-red properties of metallic films 593
Injection ratio of metal–semiconductor contacts, pulse method of measurement 447
Injection ratio of point contacts on Ge, measurement 17
Interatomic force law, and equation of state of solids 1049
Interferometer for measuring spatial frequency response of lens system 871
Interferometry, for measurement of comatic aberrations 1065
Ionic bombardment of germanium, effects produced by 213
Ionic conductivity of dilute KCl solutions at cm wavelengths 65
Ionic solutions, aqueous, dielectric constant 357
Ionic solutions, x-ray absorption spectra, origin of fine structure (R) 472
Ionosphere, upper, shape of irregularities in 493
Ionospheric drifts, effect of anisometric amplitude patterns in measurement 481
Iteration method for determination of relaxation spectra 97

Jet formation by explosives with lined hemispherical cavities (R) 785, corr. 1158
Joshi effect, inversion potential, and light intensity (R) 177
Junctions, p–n alloy germanium, breakdown effect in (R) 182

Lead selenide and telluride, thermoelectric and galvanomagnetic effects 35
Lead sulphide, current noise measurement at audio frequencies 898
Lead sulphide, photoconductive layers, effect of adsorbed air (L) 390
Lead sulphide, p–n junctions in, measurements 697
Lead sulphide, selenide and telluride, Hall coefficient, electrical conductivity and magneto-resistance effect 22
Lens system, spatial frequency response, interferometer for measuring 871
Light, scattering, in NaCl monocrystals 325
Light spot, modulated, use in semiconductor measurements 918
Linear behaviour, Fourier series for numerical treatment 89
Liquefier-cryostat, improved miniature (L) 320
Liquid helium, phase transitions of first and second orders 521
Liquid mixtures, mutual viscosity and diffusion in 209
Liquids, delayed boiling, and radio-carbon content (R) 686
Liquids, standard, measurement of dielectric constant 422
Liquid surface, impact of solid on, sounds 541
Lloyd mirror for measurement of coherence of light wave (R) 1148
Low temperature sources, superconducting bolometer detector of thermal radiation from 801
Low temperatures, *see also* Helium temperatures.
Low temperature specific heat of α and β magnanese 830
Luminescence, infra-red, of solid halogens (L) 53
Luminescence, *see also* Cathodo-luminescence

PAGE
Magnesium oxide, field enhanced cathodo-luminescence in (R) . . . 566
Magnetic susceptibility of metallic cerium . . . 859
Magnetic susceptibility of metallic neodymium (R) . . . 181
Magnetic viscosity, influence of temperature . . . 345
Magnetism, *see also* Domain structure.
Magnetization of high coercivity materials, thermal effects . . . 849
Magnetization processes in ferrites, and accompanying heat changes . . . 304
Magneto-resistance effect, electrical conductivity and Hall coefficient of lead sulphide, selenide and telluride . . . 22
Magnetothermal effect, separation of reversible and irreversible components . . . 642
Manganese, α and β, low temperature specific heat . . . 830
Materials, extensively cracked, apparent strength . . . 723
Metal electrodes, effect in electron bombardment conductivity . . . 620
Metallic films, infra-red properties . . . 593
Metal surfaces, tarnished, high-field electron emission from . . . 504
Metal, transfer between electrodes in high vacuum with high electric field (R) . . . 564
Metal transfer and wear process . . . 400
Metals, cold-worked, deformation faults in (R) . . . 51
Metals, recovery and recrystallization, in terms of restoration of transient creep properties . . . 257
Metals, *see also* Films.
Microscope, phase contrast, basic approximation in theory . . . 1073
Microwave refractive indices of water vapour, air, oxygen, nitrogen and argon . . . 833

Negative glow, helium screened probe measurements . . . 836
Neodymium, metallic, magnetic susceptibility (R) . . . 181
Nickel and alloys, ferromagnetic resonance . . . 713
Noise, current, in lead sulphide, measurement at audio frequencies . . . 898
Noise, excess, of Si filaments, effect of illumination (R) . . . 968
Noise from hot-cathode discharges . . . 292
Noise, semiconductor, distribution function (L) . . . 690
Non-metallic wear, experimental investigation . . . 106

Optical systems, automatic design . . . 729
Oxide films, examination by reflection electron microscopy (R) . . . 178

Particles, discrete, mean velocity in turbulent flow in pipe . . . 1095
Phase contrast, *see also* Microscope.
Phase transitions of first and second orders . . . 521
Phosphors, alkali halide, thermoluminescence . . . 249
Photoconductive layers of PbS, effect of adsorbed air (L) . . . 390
Photoconductivity, in calcium tungstate . . . 148
Photoeffects in condensed aromatic hydrocarbons (R) . . . 974
Photofluorescene of anthracene crystals, absolute quantum efficiency . . . 241
Point contacts, *see* Contacts, point, Germanium.
Polar solutions, dilute liquid, dielectric absorption in . . . 1038
Potassium chlorate, thermal expansion . . . 75, corr. 977
Powder and granular ridges in sound field, study . . . 462
Pressure, isothermal, and dielectric properties of eugenol, glycerol and water (R) . . . 790

Quantum efficiency, absolute, of photofluorescence of anthracene crystals . . . 241
Quartz, fused, radiation induced colour centres in . . . 223

Radiation, thermal, *see* Thermal radiation
Radio-carbon content and delayed boiling of liquids (R) . . . 686
Ray tracing, numerical, in electron lenses . . . 526
Rectification, diffusion theory, effect of field-dependent mobilities . . . 366
Rectifiers, germanium, point contact, and thermistors, turnover effect . . . 908

PAGE

Reflection, oblique, at cylindrical surfaces, properties of fine virtual slits formed by. 649
Refractive index determination by central illumination method 1081
Refractive indices of water vapour, air, oxygen, nitrogen and argon at 72 kMc/s 833
Relaxation spectra, iteration method for determination 97
Resonance, ferromagnetic, *see* Ferromagnetic.
Retina, human, photochemical reaction in 862

Scattering of light in NaCl monocrystals 325
Scintillation efficiency, absolute, of anthracene crystals 929
Scintillators, borazole and boron nitride (L) 253
Secondary electron emission from solids, theory 657
Semiconductor contacts, effect of field emission on behaviour 881
Semiconductor, fluctuations of number of electrons and holes 661
Semiconductor measurements, use of modulated light spot in 918
Semiconductor–metal contacts, pulse method for measuring injection ratio 447
Semiconductor noise, distribution function (L) 690
Semiconductors, Auger effect in (L) 319
Semiconductors, carrier concentration disturbances 310, (L) 793
Semiconductors, field emission from 746
Semiconductors, free carriers, and infra-red Faraday effect (R) 973
Semiconductors, measurements on p–n junctions in PbS 697
Semiconductors, minority carrier lifetimes, measurement (R) 562
Semiconductors, n-type, with acceptor traps, hole lifetime 894
Semiconductors, *see also* relevant materials.
Shock wave, plane, passage through wire gauze 1008
Silicon filaments, effect of illumination on excess noise (R) 968
Silicon, modulation of surface conductance by external electric fields. 10
Silicon, n-type, measurements of minority carrier lifetime 121
Silicon, single crystals, lamellar defects in 111
Solid, sounds of impact on liquid surface 541
Solidifying fluid, temperature changes in 277
Solids, equation of state, and interatomic force law 1049
Solids, secondary electron emission, theory 657
Sound from circular cylinders in air stream, measurement 453
Sound field, powder and granular ridges in 462
Space–time of constant curvature, Whitehead's law of gravitation in 944
Spark gaps, electron emission in, interpretation of experiments 513
Specific heat, low temperature, of α and β manganese 830
Spectra, *see* Absorption, Fluorescence, Relaxation, X-ray.
Spectrometer, γ-ray, curved crystal, precision 1117
Striations, moving, diffusion in (R) 383
Superconduction, *see also* Bolometer.
Supercooled water, freezing 193
Surface conductance of Ge, modulation by pulsed fields (R) 1154
Surface of growing deposits, temperatures at (L) 474
Surface recombination in Ge in strong electric fields 353
Surface states on germanium (R) 1152
Susceptibility, magnetic, *see* Magnetic susceptibility

Temperature dependence of dielectric constant of diamond (R) 315
Temperature, influence on magnetic viscosity 345
Thermal diffusion equation, with boundary conditions for solidification or melting of materials initially at fusion temperature 277
Thermal expansion of potassium chlorate 75, corr. 977
Thermal radiation from low temperature sources, superconducting bolometer as detector 801
Thermal vibrations of atoms in cubic crystals, *see under* Crystals.
Thermistors and point contact Ge rectifiers, turnover effect 908

PAGE
Thermoelectric and galvanomagnetic effects in lead selenide and telluride . . 35
Thermoluminescence of alkali halide phosphors 249
Thermometers, carbon resistor (R) 388
Thermometer, resistance, for use at He temperatures (R) 386
Transfer of metal between electrodes in high vacuum with high electric field (R) . 564
Turbulence and entrainment in open gas streams 234
Turbulent flow in pipe, mean velocity of discrete particles 1095

Viscosity, mutual, and diffusion in liquid mixtures 209

Water bells, differential analyser solution (L) 54
Water, supercooled, freezing 193
Water vapour, adsorption by anodized Al, electrical effects 1105
Waves, electromagnetic, *see under* Diffraction.
Waves, shock, *see* Shock waves.
Wear, non-metallic, experimental investigation 106
Wear process and metal transfer 400
Wear, relation to friction for boundary-lubricated surfaces 603

X-ray absorption spectra of ionic solutions, origin of fine structure (R) . . 472
X-ray diffraction camera, optically focusing 1054
X-ray extinction and effect of cold work on integrated intensities 577

INDEX OF AUTHORS (WITH TITLES)

PAGE
Alfrey, G. F., and Taylor, J. B. : Electroluminescence in single crystals of zinc sulphide 775
Allen, J. W., with Mackintosh, I. M. : The lifetime of added carriers in InSb 985
Andersson, B , Ryde, N : A precision curved crystal gamma-ray spectrometer 1117
Archard, G. D. : Two new simplified systems for the correction of spherical aberration in electron lenses 156
Archard, G. D. : A possible chromatic correction system for electron lenses 817
Arthur, J. B., Bardsley, W., Brown, M. A. C. S., and Gibson, A. F. : Carrier extraction in germanium 43
Arthur, J. B., Bardsley, W., Gibson, A. F., and Hogarth, C. A. : On the measurement of minority carrier lifetime in n-type silicon 121
Avery, D. G., and Gunn, J. B. : The use of a modulated light spot in semiconductor measurements 918

Baker, L. R. : An interferometer for measuring the spatial frequency response of a lens system 871
Banbury, P. C., and Houghton, J. : Measurement of injection ratio of point contacts on germanium 17
Barber, D. : Measurements of current noise in lead sulphide at audio frequencies 898
Bardsley, W., with Arthur, J. B., Brown, M. A. C. S., and Gibson, A. F. : Carrier extraction in germanium 43
Bardsley, W., with Arthur, J. B., Gibson, A. F., and Hogarth, C. A. : On the measurement of minority carrier lifetime in n-type silicon 121
Bardsley, W., with Hogarth, C. A. : The effect of illumination on the excess noise of silicon filaments (R) 968
Batchelor, G. K., Binnie, A. M., and Phillips, O. M. : The mean velocity of discrete particles in turbulent flow in a pipe 1095
Bates, L. F., Leach, S. J., and Loasby, R. G. : The magnetic susceptibility of metallic cerium 859
Bates, L. F., Leach, S. J., Loasby, R. G., and Stevens, K. W. H. : The magnetic susceptibility of metallic neodymium (R) 181
Bates, L. F., and Martin, D. H. : A study of domain structures in alnico 537
Bates, L. F., and Sherry, N. P. R. : The heat changes accompanying magnetization processes in ferrites 304
Bates, L. F., and Sherry, N. P. R. : The direct separation of the reversible and irreversible components of the magnetothermal effect 642
Bates, L. F., and Simpson, A. W. : The thermal effects associated with the magnetization of high-coercivity materials 849
Bell, D. A. : Distribution function of semiconductor noise (L) 690
Binnie, A. M., with Batchelor, G. K., and Phillips, O. M. : The mean velocity of discrete particles in turbulent flow in a pipe 1095
Black, G. : On the automatic design of optical systems 729
Booth, G. L., Hoare, F. E., and Murphy, B. T. : The low temperature (11°–20°K) specific heat of α and β manganese 830
Bose, H. N. : Thermoluminescence of alkali halide phosphors 249
Bowman, J. E., and Niblett, G. B. F., The passage of a plane shock wave through a wire gauze 1008
Brindley, G. S. : A photochemical reaction in the human retina 862
Brodie, I. : D.C. and pulsed emission from L-cathodes (R) 1146
Brown, G. Burniston : A theory of action-at-a-distance 672
Brown, M. A. C. S., with Arthur, J. B., Bardsley, W., and Gibson, A. F. : Carrier extraction in germanium 43
Browne, P. F. : The transfer of metal between electrodes in a high vacuum on application of a high electric field (R) 564

PAGE
Burgess, R. E.: Low frequency noise in illuminated germanium diodes (R) . . 569
Burgess, R. E.: Fluctuations of the numbers of electrons and holes in a semiconductor 661
Burgess, R. E.: The a.c. admittance of temperature-dependent circuit elements 766
Burgess, R. E.: Carrier concentration disturbances in semiconductors (L) . . 793
Burgess, R. E.: The turnover phenomenon in thermistors and in point-contact germanium rectifiers 908

Cade, R.: An electrostatic problem involving a non-linear fluid dielectric . . 1
Cartz, L.: Thermal vibrations of atoms in cubic crystals: I—The temperature variation of thermal diffuse scattering of x-rays by lead single crystals . . 951
Cartz, L.: Thermal vibrations of atoms in cubic crystals: II—The amplitude of atomic vibrations 957
Chaudhri, R. M.: A mercury safety valve for electrolysis of heavy water (R) . . 1157
Churchman, A. T., with Geach, G. A., and Franks, J.: Lamellar defects in single crystals of silicon 111
Clarke, W. W. H., and Jacob, L.: Some anomalous emission effects in low energy electron beams 805
Colson, I. D., with Knott, R. D., and Young, M. R. P.: Breakdown effect in p–n alloy germanium junctions (R) 182
Cook, J. R.: Photoconductivity in calcium tungstate 148
Cooke, A. H., Whitley, S., and Wolf, W. P.: Demagnetization experiments on cerium ethyl sulphate 415
Crompton, R. W., with Huxley, L. G. H.: A note on the diffusion in a gas of electrons from a small source (R) 381
Cussins, W. D.: Effects produced by the ionic bombardment of germanium . . 213

Dayal, B., and Sharma, R. S.: Inter-atomic force law and equation of state of solids . 1049
Deland, E. C., with Lance, G. N.: On the differential analyser solution of the water bells problem (L) 54
Dodd, C.: Radio-carbon content and delayed boiling of liquids 686
Dumbleton, M. J.: The infra-red luminescence of solid halogens (L) . . . 53
Durrant, N. F.: Measurement of minority carrier lifetimes in semiconductors (R) . 562

Emara, Sayeda H., with Tolansky, S.: An occasional mode of growth in diamond . 559
Essen, L., with Hartshorn, L., and Parry, J. V. L.: The measurement of the dielectric constant of standard liquids 422

Fairweather, A.: Dielectric absorption in dilute, liquid, polar solutions: a new approach 1038
Farris, V. D.: Diffusion in moving striations (R) 383
Farvis, W. E. J., with Pringle, D. H.: Screened probe measurements in the helium negative glow 836
Faust, R. C.: Refractive index determinations by the central illumination (Becke line) method 1081
Firmin, J. C., and Oatley, C. W.: Effects produced by metal electrodes in studies of electron bombardment conductivity 620
Fowler, R. G.: Constriction of the positive column of a glow discharge at high pressure 130
Francis, G.: The growth of electrodeless discharges in hydrogen 137
Francis, G.: The growth of an electron avalanche retarded by its own spaec charge . 369
Franks, A.: An optically focusing x-ray diffraction camera 1054
Franks, J., Geach, G. A., and Churchman, A. T.: Lamellar defects in single crystals of silicon 111
Fröhlich, H.: Theory of secondary electron emission from solids 657
Froome, K. D.: The refractive indices of water vapour, air, oxygen, nitrogen and argon at 72 kMc/s 833

PAGE

Garlick, G. F. J., with Halperin, A. : The absorption spectrum of excited crystals of cadmium sulphide 758
Gates, J. W. : The measurement of comatic aberrations by interferometry . . 1065
Geach, G. A., with Franks, J., and Churchman, A. T. : Lamellar defects in single crystals of silicon 111
Gerlich, D., with Many, A., and Harnik, E. : Some measurements connected with carrier deficit phenomena in germanium (R) 970
Gerrard, J. H. : Measurements of the sound from circular cylinders in an air stream. 453
Gerson, N. C. : Diurnal variation in auroral activity 408
Gibson, A. F., with Arthur, J. B., Bardsley, W., and Brown, M. A. C. S. : Carrier extraction in germanium 43
Gibson, A. F., with Arthur, J. B., Bardsley, W., and Hogarth, C. A. : On the measurement of minority carrier lifetime in n-type silicon 121
Gough, R. H. : Entrainment and turbulence in open gas streams 234
Greenough, G. B., and Smith, E. M. : Deformation faults in cold-worked metals . 51
Greenwood, J. A., and Tabor, D. : Deformation properties of friction junctions . 609
Guggenheim, E. A. : The computation of electric dipole moments (L) . 186 corr. 253
Gunn, J. B., with Avery, D. G. : The use of a modulated light spot in semiconductor measurements 918

Hall, B. I. H. : Experimental investigation of the motions of electrons in a gas in the presence of a magnetic field 334
Hall, E. O., with Street, R. : Magnetic domain structure in twinned crystals of silicon iron 1033
Halliday, J. S., and Hirst, W. : The examination of oxide films by reflection electron microscopy (R) 178
Halperin, A., and Garlick, G. F. J. : The absorption spectrum of excited crystals of cadmium sulphide 758
Harnik, E., with Many, A., and Gerlich, D. : Some measurements connected with carrier deficit phenomena in germanium (R) 970
Hartshorn, L., Parry, J. V. L., and Essen, L. : The measurement of the dielectric constant of standard liquids 422
Harwood, M. G. : The crystal structure of lanthanum–strontium manganites . . 586
Henisch, H. K., and Reynolds, W. N. : Surface recombination in germanium in the presence of strong electric fields 353
Hill, N. E. : Mutual viscosity and diffusion in liquid mixtures 209
Hirsch, P. B., Kelly, A., and Menter, J. W. : The structure of cold worked gold : I—a study of electron diffraction 1132
Hirst, W., with Halliday, J. S.: The examination of oxide films by reflection electron microscopy (R) 178
Hoare, F. E., Matthews, J. C., and Yates, B. : Carbon resistor thermometers (R) . 388
Hoare, F. E., with Booth, G. L., and Murphy, B. T. : the low-temperature (11°–20°K) specific heat of α and β manganese 830
Hodgson, J. N. : The infra-red properties of some metallic films 593
Hogarth, C. A., and Bardsley, W. : The effect of illumination on the excess noise of silicon filaments (R) 968
Hogarth, C. A., with Arthur, J. B., Bardsley, W., and Gibson, A. F. : On the measurement of minority carrier lifetime in n-type silicon 121
Houghton, J., with Banbury, P. C. : Measurement of injection ratio of point contacts on germanium 17
Hughes, G., and Spurr, R. T. : An experimental investigation of non-metallic wear. 106
Hulbert, J. A., and Jones, G. O. : The superconducting bolometer as a detector of thermal radiation from low-temperature sources 801
Hulme, K. F. : On the distribution of impurity in crystals grown from impure unstirred melts 393
Humphreys-Owen, S. P. F. : The scattering of light in sodium chloride monocrystals 325

PAGE
Huntington, H. B. : Modification of the Peierls–Nabarro model for edge dislocation core 1043
Hurst, C. : Phase transitions of the first and second orders 521
Huxley, L. G. H., and Crompton, R. W. : A note on the diffusion in a gas of electrons from a small source (R) 381

Jacob, L., with Clarke, W. W. H. : Some anomalous emission effects in low energy electron beams 805
Jason, A. C., and Wood, J. L. : Some electrical effects of the adsorption of water vapour by anodized aluminium 1105
Jennings, J. C. E., and Pratt, R. G. : Numerical ray tracing in electron lenses . 526
Jones, D. A. : Growth of lead fluoride crystals from the melt. 165
Jones, G. O. : An improved miniature liquefier-cryostat (L). 320
Jones, G. O., with Hulbert, J. A. : The superconducting bolometer as a detector of thermal radiation from low-temperature sources 801

Kahn, F. D. : On the basic approximation in the theory of the phase contrast microscope 1073
Kelly, A., with Hirsch, P. B., and Menter, J. W. : The structure of cold worked gold: I—a study of electron diffraction 1132
Kennedy, A. J. : Recovery and recrystallization in metals examined in terms of the restoration of the transient creep properties. 257
Kerridge, M. : Metal transfer and the wear process 400
Kirkbride, J. : Borazole and boron nitride scintillators (L) 253
de Klerk, J., and Musgrave, M. J. P. : Internal conical refraction of transverse elastic waves in a cubic crystal 81
Knott, R. D., Colson I. D., and Young, M. R. P. : Breakdown effect in p–n alloy germanium junctions (R) 182
Kreith, F., and Romie, F. E. : A study of the thermal diffusion equation with boundary conditions corresponding to solidification or melting of materials initially at the fusion temperature 277

Lance, G. N., and Deland, E. C. : On the differential analyser solution of the water bells problem (L) 54
Landsberg, P. T. : The effect of field-dependent mobilities on the diffusion theory of rectification 366
Leach, S. J., with Bates, L. F., and Loasby, R. G. : The magnetic susceptibility of metallic cerium 859
Leach, S. J., with Bates, L. F., Loasby, R. G., and Stevens, K. W. H. : The magnetic susceptibility of metallic neodymium (R) 181
Leak, G. M., with Thomas, W. R. : The binding energy of nitrogen in a dislocation . 1001
Lees, J., and Walton, S. : The influence of pressure on metal–germanium contacts . 922
Lees, J., and Walton, S. : Surface states on germanium (R) 1152
Levy, M., and Varley, J. H. O.: Radiation induced colour centres in fused quartz . 223
Lewis, T. J. : The mechanism of high-field electron emission from tarnished metal surfaces 504
Lewis, T. J. : Some factors influencing field emission and the Fowler–Nordheim law 938
Liebmann, G. : The field distribution in asymmetrical magnetic electron lenses (R) . 679
Liebmann, G. : The magnetic pinhole electron lens (R) 682
Liebmann, G. : A unified representation of magnetic electron lens properties . . 737
Little, V. I. : The dielectric constant of aqueous ionic solutions 357
Little, V. I., and Smith, V. : The ionic conductivity of dilute potassium chloride solutions at centimetric wavelengths 65
Loasby, R. G., with Bates, L. F., and Leach, S. J. : The magnetic susceptibility of metallic cerium 859
Loasby, R. G., with Bates, L. F., Leach, S. J., and Stevens, K. W. H. : The magnetic susceptibility of metallic neodymium (R) 181
Lonappan, M. A. : Thermal expansion of potassium chlorate . . . 75, corr. 977

PAGE

Low, G. G. E. : Modulation of the surface conductance of germanium and silicon by external electric fields . . . 10
Low, G. G. E. : Carrier concentration disturbances in semiconductors . . . 310
Low, G. G. E. : A pulse method for measuring the injection ratio of metal–semiconductor contacts . . . 447
Low, G. G. E.: Modulation of the surface conductance of germanium by pulsed electric fields (R) . . . 1154

McIlraith, A. H. : The properties of fine virtual slits formed by oblique reflection at cylindrical surfaces . . . 649
McIlraith, A. H. : The measurement of the coherence of a light wave by means of a Lloyd mirror (R) . . . 1148
Mackintosh, I. M., and Allen, J. W. : The lifetime of added carriers in InSb . . . 985
Many, A., Gerlich, D., and Harnik, E. : Some measurements connected with carrier deficit phenomena in germanium (R) . . . 970
Martin, D. H., with Bates, L. F. : A study of domain structures in alnico . . . 537
Matthews, J. C., with Hoare, F. E., and Yates, B. : Carbon resistor thermometers (R) 388
Menter, J. W., with Hirsch, P. B., and Kelly, A. : The structure of cold worked gold: I—a study of electron diffraction . . . 1132
Millard, D. J., Newman, P. C., and Phillips, J. W. : The apparent strength of extensively cracked materials . . . 723
Mitchell, E. W. J. : The infra-red Faraday effect due to free carriers in a semiconductor (R) . . . 973
Mohanty, S. R., and Rao, P. Srinivasa : Inversion potential of the Joshi effect and light intensity (R) . . . 177
Morant, M. J. : Interpretation of experiments on electron emission in spark gaps . 513
Moss, T. S. : Measurements on p–n junctions in lead sulphide . . . 697
Mossop, S. C. : The freezing of supercooled water . . . 193
Murphy, B. T., with Booth, G. L., and Hoare, F. E. : The low temperature (11°–20°K) specific heat of α and β manganese . . . 830
Musgrave, M. J. P., with de Klerk, J. : Internal conical refraction of transverse elastic waves in a cubic crystal . . . 81

Narasimhan, P. T. : Temperature dependence of the dielectric constant of diamond (R) . . . 315
Newman, P. C., with Millard, D. J., and Phillips, J. W. : The apparent strength of extensively cracked materials . . . 723
Niblett, G. B. F., with Bowman, J. .E : The passage of a plane shock wave through a wire gauze . . . 1008
Nigam, A. N. : The origin of fine structures in X-ray absorption spectra of ionic solutions . . . 472
Northrop, D. C., and Simpson, O. : Photoeffects in condensed aromatic hydrocarbons (R) . . . 974

Oatley, C. W., and Firmin, J. C. : Effects produced by metal electrodes in studies of electron bombardment conductivity . . . 620

Parkinson, D. H., and Roberts, L. M. : A resistance thermometer for use at helium temperatures (R) . . . 386
Parry, J. V. L., with Hartshorn, L., and Essen, L. : The measurement of the dielectric constant of standard liquids . . . 422
Phillips, G. J., and Spencer, M. : The effects of anisometric amplitude patterns in the measurement of ionospheric drifts . . . 481
Phillips, J. H., Woolley, J. C., and Street, R. : The influence of temperature on magnetic viscosity . . . 345
Phillips, J. W., with Millard, D. J., and Newman, P. C. : The apparent strength of extensively cracked materials . . . 723

PAGE
Phillips, O. M., with Batchelor, G. K., and Binnie, A. M.: The mean velocity of discrete particles in turbulent flow in a pipe 1095
Pincherle, L.: Auger effect in semiconductors (L) 319
Plessner, K. W., and West, R.: Replacement of Ti in $BaTiO_3$ ceramic by Si and Ge (R) 1150
Potok, M. H. N.: The diffraction of a non-monochromatic electromagnetic wave by a slit and a grating 171
Pratt, R. G., with Jennings, J. C. E.: Numerical ray tracing in electron lenses 526
Pratt, R. G., with Smollett, M.: The effect of adsorbed air on photoconductive layers of lead sulphide (L) 390
Pringle, D. H., and Farvis, W. E. J.: Screened probe measurements in the helium negative glow 836
Putley, E. H.: The Hall coefficient, electrical conductivity and magneto-resistance effect of lead sulphide, selenide and telluride 22
Putley, E. H.: Thermoelectric and galvanomagnetic effects in lead selenide and telluride 35

Rabinowicz, E.: The relation between friction and wear for boundary-lubricated surfaces 603
Rao, P. Srinivasa, with Mohanty, S. R.: Inversion potential of the Joshi effect and light intensity (R) 177
Rayner, C. B.: Whitehead's law of graviation in a space–time of constant curvature 944
Reich, K. H., with Standley, K. J.: Ferromagnetic resonance in nickel and in some of its alloys 713
Reynolds, W. N., with Henisch, H. K.: Surface recombination in germanium in the presence of strong electric fields 353
Richardson, E. G.: The sounds of impact of a solid on a liquid surface . . 541
Roberts, F. F., and Tillman, J. R.: The barrier height of point-contact germanium diodes inferred from measurements of the voltage dependence of capacitance (R) 113
Roberts, L. M., with Parkinson, D. H.: A resistance thermometer for use at helium temperatures (R) 386
Roesler, F. C.: Some applications of Fourier series in the numerical treatment of linear behaviour 89
Roesler, F. C., and Twyman, W. A.: An iteration method for the determination of relaxation spectra 97
Romie, F. E., with Kreith, F.: A study of the thermal diffusion equation with boundary conditions corresponding to solidification or melting of materials initially at the fusion temperature 277
Rose, F. W. G., and Sandiford, D. J.: An easy derivation of the hole lifetime in an n-type semiconductor with acceptor traps 894
Ryde, N., and Andersson, B.: A precision curved crystal gamma-ray spectrometer 1117

Sandiford, D. J., with Rose, F. W. G.: An easy derivation of the hole lifetime in an n-type semiconductor with acceptor traps 894
Scaife, B. K. P.: Isothermal pressure dependence of the dielectric properties of eugenol, glycerol and water (R) 790
Sharma, R. S., with Dayal, B.: Inter-atomic force law and equation of state of solids 1049
Sherry, N. P. R., with Bates, L. F.: The heat changes accompanying magnetization processes in ferrites 304
Sherry, N. P. R., with Bates, L. F.: The direct separation of the reversible and irreversible components of the magnetothermal effect 642
Sillars, R. W.: The effect of field emission on the behaviour of semiconductor contacts 881
Simpson, A. W., with Bates, L. F.: The thermal effects associated with the magnetization of high-coercivity materials 849
Simpson, O., with Northrop, D. C.: Photoeffects in condensed aromatic hydrocarbons (R). 974
Singh, Sampooran: On the jet formation by explosives with lined hermispherical cavities (R) . . . . 785, corr. 1158

PAGE

Smallman, R. E., with Williamson, G. K.: X-ray extinction and the effect of cold work on integrated intensities . . . 577
Smith, E. M., with Greenough, G. B.: Deformation faults in cold-worked metals (R) 51
Smith, V., with Little, V. I.: The ionic conductivity of dilute potassium chloride solutions at centimetric wavelengths . . . 65
Smollett, M., and Pratt, R. G.: The effect of adsorbed air on photoconductive layers of lead sulphide (L) . . . 390
Spear, W. E.: Electron bombardment effects in thin dielectric layers . . . 991
Spencer, M.: The shape of irregularities in the upper ionosphere . . . 493
Spencer, M., with Phillips, G. J.: The effects of anisometric amplitude patterns in the measurement of ionospheric drifts . . . 481
Spurr, R. T., with Hughes, G.: An experimental investigation of non-metallic wear 106
Standley, K. J., and Reich, K. H.: Ferromagnetic resonance in nickel and in some of its alloys . . . 713
Stevens, K. W. H., with Bates, L. F., Leach, S. J., and Loasby, R. G.: The magnetic susceptibility of metallic neodymium (R) . . . 181
Stratton, R.: Field emission from semiconductors . . . 746
Street, R., and Hall, E. O.: Magnetic domain structure in twinned crystals of silicon iron . . . 1033
Street, R., with Phillips, J. H., and Woolley, J. C.: The influence of temperature on magnetic viscosity . . . 345

Tabor, D., with Greenwood, J. A.: Deformation properties of friction junctions . 609
Taylor, J. B., with Alfrey, G. F.: Electroluminescence in single crystals of zinc sulphide . . . 775
Tebble, R. S.: The Barkhausen effect . . . 1017
Thomas, W. R., and Leak, G. M.: The binding energy of nitrogen in a dislocation 1001
Thong Saw Pak: Noise from hot-cathode discharges. . . . 292
Tillman, J. R., with Roberts, F. F.: The barrier height of point-contact germanium diodes inferred from measurements of the voltage dependence of capacitance (R) 113
Tolansky, S., and Emara, Sayeda H.: An occasional mode of growth in diamond . 559
Tolansky, S., and Williams, Anne P.: Directional hardness studies on single crystals of tin and bismuth . . . 548
Twyman, W. A., with Roesler, F. C.: An iteration method for the determination of relaxation spectra . . . 97

Varley, J. H. O., with Levy, M.: Radiation induced colour centres in fused quartz 223

Waller, M. D.: A study of powder and granular ridges in a sound field . . . 462
Walton, S., with Lees, J.: The influence of pressure on metal–germanium contacts 922
Walton, S., with Lees, J.: Surface states on germanium . . . 1152
West, R., with Plessner, K. W.: Replacement of Ti in BaTiO$_3$ ceramic by Si and Ge (R) . . . 1150
Whitley, S., with Cooke, A. H., and Wolf, W. P.: Demagnetization experiments on cerium ethyl sulphate . . . 415
Williams, Anne, P., with Tolansky, S.: Directional hardness studies on single crystals of tin and bismuth . . . 548
Williamson, G. K., and Smallman, R. E.: X-ray extinction and the effect of cold work on integrated intensities . . . 577
Wilman, H.: The temperature attained at the surfaces of gowing deposits, and the origin of stress in deposits (L) . . . 474
Wilson, R. W.: The contact resistance and mechanical properties of surface films on metals . . . 625
Wolf, W. P., with Cook, A. H., and Whitley, S.: Demagnetization experiments on cerium ethyl sulphate . . . 415
Wood, J. L., with Jason, A. C.: Some electrical effects of the adsorption of water vapour by anodized aluminium . . . 1105

PAGE
Woods, J., and Wright, D. A.: Field enhanced cathodo-luminescence in magnesium oxide (R) 566
Woolley, J. C., with Phillips, J. H., and Street, R.: The influence of temperature on magnetic viscosity 345
Wright, D. A., with Woods, J.: Field enhanced cathodo-luminescence in magnesium oxide (R) 566
Wright, D. J.: Hysteresis of the angle of contact of mercury against steel 297
Wright, G. T.: Absolute quantum efficiency of photofluorescence of anthracene crystals 241
Wright, G. T.: Fluorescence excitation spectra and quantum efficiencies of organic crystals 701
Wright, G. T.: Absolute scintillation efficiency of anthracene crystals 929

Yates, B., with Hoare, F. E., and Matthews, J. C.: Carbon resistor thermometers (R) 388
Young, M. R. P., with Knott, R. D., and Colson, I. D.: Breakdown effect in p–n alloy germanium junctions (R) 182

INDEX TO REVIEWS OF BOOKS

PAGE
Aeronautial Research Council, Compressible Flow Tables Panel: *A Selection of Graphs for use in Calculations of Compressible Airflow* 59
Alfvén, H.: *On the Origin of the Solar System* 693
Allen, D. N. de G.: *Relaxation Methods* 190
Adirowitsch, E. I.: *Some Problems in the Theory of the Luminescence of Crystals* 117

Bergmann, L.: *Der Ultrashcall and seine Anwendung in Wissenschaft und Technik* 116
Beteille, R.: *Elements de base des mesures en vol* 575
Booth, A. D.: *Numerical Methods* 980
Born, M., and Huang, K.: *Dynamical Theory of Crystal Lattices* 978
Bowden, B. V. (Ed.): *Faster than Thought* 189
Boyd, R. L. F., and Seaton, M. J.: *Rocket Exploration of the Upper Atmosphere* 254
Brown, T. B.: *Electronics* 983

Chandrasekhar, S., and Elbert, D. D.: *The Illumination and Polarisation of the Sunlit Sky on Rayleigh Scattering* 693
Constant, F. W.: *Theoretical Physics—Mechanics of Particles, Rigid and Elastic Bodies, Fluids and Heat Flow* 60

Ditchburn, R. W.: *The Students' Physics*, Vol. 1.—*Light* 797

Ferraro, V. C. A.: *Electromagnetic Theory* 57
Françon, M.: *Contraste de phase et constrate par interferences* 56
Fretter, W. B.: *Introduction to Experimental Physics* 796

Goudet, G.: *Les fonctions de Bessel* 794
Green, A. E., and Zerna, W.: *Theoretical Elasticity* 254
Guggenheim, E. A., and Prue, J. E.: *Physicochemical Calculations* 692

Hoel, P. G.: *Introduction to Mathematical Statistics* 795

PAGE
Ingersoll, L. R., Zobell, O. J., and Ingersoll, A. C. : *Heat Conduction with Engineering, Geological and other Applications* 57
Institute of Petroleum : *Applied Mass Spectrometry* 574
Institute of Physics : *Physics of Particle Size Analysis* 477

Johns, R. V., and Ware, W. F. : *Advanced Level Examples in Physics* 478

Kronig, R. (Ed.) : *Textbook of Physics* 61
Kuiper, G. P. : *The Sun* 188

Lense, J. : *Reihenentwicklungen in der mathematischen Physik* 696
Lewis, I. A. D., and Wells, F. H. : *Millimicrosecond Pulse Techniques* 981
Lichnerowicz, A. : *Théories relativistes de la gravitation et de l'électromagnétisme* . 979
Linfoot, E. H. : *Recent Advances in Optics* 798
Lösch, F., *Siebenstellige Tafeln der elementaren transzendenten Funktionen* . . 1159
Lovell, A. C. B., and Colleagues : *Meteor Astronomy* 118

Marton, L. : *Advances in Electronics*, Vol. V 59
Marton, L. : *Advances in Electronics and Electron Physics*, Vol. VI 573
Morey, G. W. : *The Properties of Glass* 692

Payne, B. O. : *Microscope Design and Construction* 574
Pohl, R. W. : *Einführung in die Physik, Band 3, Optik und Atomphysik* . . . 116

Reiner, M. : *Building Materials, their Elasticity and Inelasticity* 572
Riabouchinsky, D. P. : *Mémories sur la mécanique des fluides* 56
Royal Meteorological Society : *Proceedings of the Toronto Meteorological Conference*, 1953 62
Rosenhead, L. : *See* Aeronatual Research Council

Spenke, E. L. : *Elektronische Halbleiter—eine Einführung in die Physik der Gleichrichter und Transistoren* 694
Starkey, B. J. : *Laplace Transforms for Electrical Engineers* 794
Starr, A. T. : *Electronics* 58
Struik, D. J. : *A Concise History of Mathematics* 322

U.S. Department of Commerce, N.B.S. Applied Mathematics Series 32 : *Table of Sine and Cosine Integrals for Arguments from* 10 *to* 100 322
U.S. Department of Commerce, N.B.S. Applied Mathematics Series 34 : *Table of the Gamma Function for Complex Arguments* 572

Ziller, A. : *Methodes de différentiation et d'integration numeriques* 978
Zwikker, C. : *Physical Properties of Solid Materials* 60

F